现代数学基础丛书

数学规划导论



科学出版社

现 代 数 学 基 础 丛 书

数学规划导论

徐增堃 著

浙江师范大学首批重点建设教材项目
浙江师范大学科研出版基金资助项目
浙 江 省 重 点 学 科 基 金 资 助

科 学 出 版 社

2001

内 容 简 介

本书分线性规划与非线性规划两部分. 线性规划部分主要包括单纯形法、对偶理论和摄动理论. 非线性规划部分主要包括最优性条件、对偶理论和鞍点理论, 以及约束和无约束最优化中的一些重要方法. 本书注重用最基本的理论和工具推导数学规划中的重要结果, 同时力图反映本学科的某些科研发展动向.

本书适合于高等院校的学生、教师、研究生作参考书、教材或自学用书. 本书还可供有关的科学工作者参考.

图书在版编目(CIP)数据

数学规划导论/徐增堃著. -北京：科学出版社，2000.6

(现代数学基础丛书)

ISBN 7-03-008200-1

Ⅰ.数… Ⅱ.徐… Ⅲ.数学规划 Ⅳ.O221

中国版本图书馆 CIP 数据核字（1999）第 74211 号

科学出版社 出版

北京东黄城根北街 16 号

邮政编码:100717

新蕾印刷厂印刷

科学出版社发行 各地新华书店经销

*

2000 年 6 月第 一 版 开本:850×1168 1/32

2001 年 3 月第二次印刷 印张:7

印数:3 001—6 000 字数:180 000

定价: 15.00 元

(如有印装质量问题,我社负责调换〈北燕〉)

前　言

本书是为略知数学规划或对之不甚知晓但数学基础较好的人所编写的，取材难免会受到作者的偏好的影响．尽管如此，在编写的过程中还是比较注意到材料的基础性、系统性和连贯性；注意吸收当今的，特别是作者自己的研究成果和学习心得；注意指出产生结果的本质所在以及各章节之间的相互联系；同时还注意到有关理论的应用性；希望用较小的篇幅来包含尽可能多的思想和内容，以使读者对进一步的学习和研究产生兴趣或有所启迪．

徐增堃

目　　录

第一部分　线性规划

第一章　基本概念和基本性质

§1.1　引言

线性规划的中心论题是求线性函数在线性等式或不等式约束下达最小或最大值的问题.

试看一个简单的例子.

例 1.1　某厂生产甲、乙两种同类产品：

单位产品需量（原料＼产品）	甲	乙	原料可供应量
第一种原料	1	1	3500
第二种原料	1	0	1500
第三种原料	5	2	10000
单位产品利润	5	3	

问如何安排生产使总利润最多？

设生产甲、乙各为 x_1, x_2 件，则

$$
\begin{aligned}
\max\quad & 5x_1+3x_2 \\
\text{s.t.}\quad & x_1+x_2\leqslant 3500 \\
& x_1 \leqslant 1500 \\
& 5x_1+2x_2\leqslant 10000 \\
& x_1\ , x_2\geqslant 0
\end{aligned}
$$

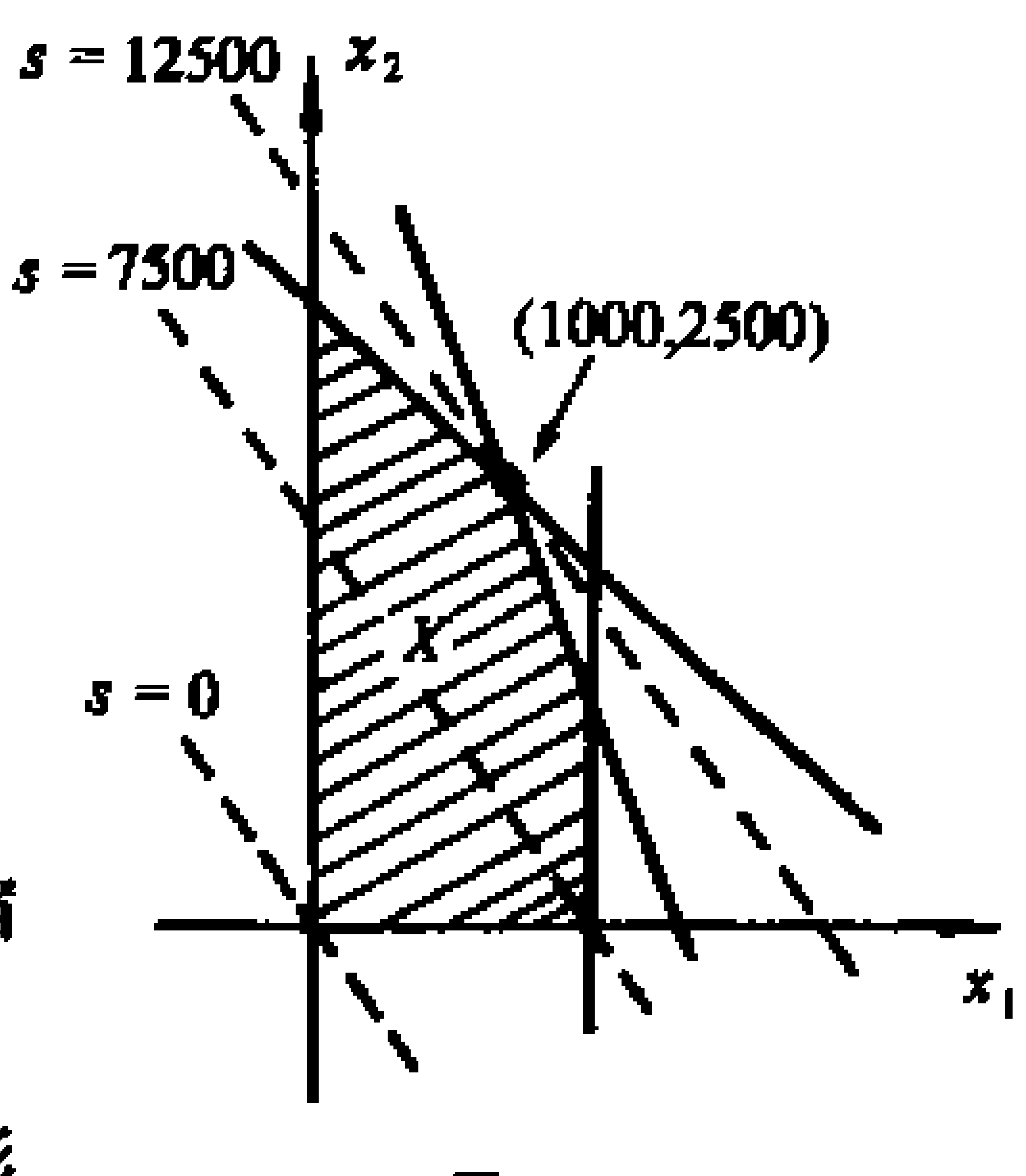


图 1.1

其中“s. t.”为英文“subject to”的缩写，表示“受限制于”.

从图 1.1 中可以看出(其中阴影

部分表示变量的取值范围，虚线是 $s=5x_1+3x_2$ 的等高线），s 的值从零连续变化增加到 12500. 因此，顶点(1000，2500)是最优解，最大利润 $s=12500$.

线性规划最基本的性质是在顶点达到极值. 通过代数的方法描述高维空间"有关图形"的顶点，然后使用经济的方法求出达到极值的顶点，这是线性规划的重要内容.

1947 年，G. B. Dantzig 提出了解线性规划的单纯形法，线性规划这门学科开始形成并迅速发展. 这主要是由于实践中的大量优化问题可以归结为线性规划. 另一方面也由于它的方法的有效性.

关于线性规划的综合材料，可参见文献[4，11，24，25].

§1.2 线性规划的基本概念

1. 标准型

线性规划的标准型如下：

$$
\text{(LP)}\qquad
\begin{aligned}
\min\ & c_1x_1+c_2x_2+\cdots+c_nx_n \\
\text{s.t.}\ & a_{11}x_1+a_{12}x_2+\cdots+a_{1n}x_n=b_1 \\
& a_{21}x_1+a_{22}x_2+\cdots+a_{2n}x_n=b_2 \\
& \cdots\qquad \cdots\qquad \cdots\qquad \cdots \\
& a_{m1}x_1+a_{m2}x_2+\cdots+a_{mn}x_n=b_m \\
& x_1\geqslant 0,x_2\geqslant 0,\cdots,x_n\geqslant 0
\end{aligned}
$$

用紧凑的矩阵、向量记号，(LP)就是

$$
\begin{aligned}
\min\ & c^Tx \\
\text{s.t.}\ & Ax=b \\
& x\geqslant 0
\end{aligned}
$$

其中 c 和 x 为 n 维列向量，b 为 m 维列向量，A 为 $m\times n$ 阶阵，$x\geqslant 0$ 表示 x 的每个分量均非负.

记 $A=(a_1,a_2,\cdots,a_n)$，其中 a_j 表 A 的第 j 列，则(LP)亦可表为

$$\min c^T x$$

$$\text{s.t.} \sum_{j=1}^{n} x_j a_j = b$$

$$x_j \geqslant 0,\ j=1,2,\cdots,n$$

在讨论中，我们视方便选用以上任一种形式，并称 $c^T x$ 为目标函数，称 $Ax=b$，$x\geqslant 0$ 为约束，称

$$X = \{x: Ax = b, x \geqslant 0\} \tag{1.1}$$

为(LP)的可行集. 由此，问题还可以描述为

$$\min_{x\in X} c^T x$$

对于标准型，若无特别声明，则约定以下假设成立：

(i) $m<n$；

(ii) A 的秩为 m；

(iii) $b\geqslant 0$.

2. 转化为标准型

我们能够指出，任何一种线性规划问题皆可等价地转化为标准型. 因此，在一般的讨论中我们只考虑标准型. 以下只列出一些常见的情况.

情况一

$$\begin{aligned} &\min c^T x \\ &\text{s.t. } Ax \leqslant b \\ &\quad\quad x \geqslant 0 \end{aligned}$$

引进松弛变量 $y=(y_1,\cdots,y_m)^T$，可得等价的

$$\begin{aligned} &\min c^T x \\ &\text{s.t. } Ax+Iy=b \\ &\quad\quad x\geqslant 0, y\geqslant 0 \end{aligned}$$

对应于标准型的阵为 (A,I_m)，对应的目标函数系数为 $\begin{pmatrix} c \\ 0 \end{pmatrix} \in R^{n+m}$，变量为 $m+n$ 个.

情况二 若情况一中有 $Ax\geqslant b$，其余不变，则有

$$\min c^T x$$
$$\text{s.t. } Ax - Iy = b$$
$$x \geqslant 0, y \geqslant 0$$

上面的 y 称为剩余变量.

情况三 如果在某个线性规划问题中,对一部分变量不要求非负约束外,它是用标准形式给出的.例如不限 $x_1 \geqslant 0$,此时的 x_1 称为自由变量.可以记

$$x_1 = u_1 - v_1, \; u_1, v_1 \geqslant 0$$

并且相应于(LP),目标函数 $c^T x$ 及等式约束中的 x_1 均以 $u_1 - v_1$ 代入,非负约束成为 $u_1 \geqslant 0, v_1 \geqslant 0, x_2, \cdots, x_n \geqslant 0$,变量个数为 $n+1$.

情况四 目标函数求极大的情况

$$\max c^T x$$
$$\text{s.t. } Ax = b$$
$$x \geqslant 0$$

可以化为等价的

$$\min -c^T x$$
$$\text{s.t. } Ax = b$$
$$x \geqslant 0$$

例 1.2 例 1.1 中的规划可利用情况一、四化为等价的

$$\min -5x_1 - 3x_2$$
$$\begin{aligned} \text{s.t. } \quad & x_1 + x_2 + x_3 = 3500 \\ & x_1 + x_4 = 1500 \\ & 5x_1 + 2x_2 + x_5 = 10000 \\ & x_i \geqslant 0, \; i = 1,2,3,4,5 \end{aligned}$$

3. 基本可行解

前节中,我们考察到线性规划的最优解在顶点处达到.为从代数的角度来研究它,我们引入下述概念.

考虑方程组

$$Ax = b \tag{1.2}$$

其中 A 是秩为 m 的 $m\times n$ 阶阵. 令 B 表由 A 的列组成的任意 $m\times m$阶非异子阵,对应的 x 的分量记为 x_B. 即有下标 $j_1,j_2,\cdots,j_m$, $1\leqslant j_i\leqslant n$, 使 $B=(a_{j_1},a_{j_2},\cdots,a_{j_m})$非异,且记 $x_B=(x_{j_1},x_{j_2},\cdots,x_{j_m})$. 则可由 $Bx_B=b$ 唯一地解出 $x_B=B^{-1}b$.

定义 1.3 令 B 表示由 A 的列所组成的任意 $m\times m$ 阶非异子阵,令不与 B 对应的 $n-m$ 个 x 的分量都为零. 这样得到的方程组(1.2)的解,称为该方程组关于基 B 的基本解. 对应于 B 的各列的 x 的分量称为基本变量,其余的分量称为非基本变量.

现在,我们在(1.2)中加入非负约束,得

$$\begin{cases} Ax=b \\ x\geqslant 0 \end{cases} \tag{1.3}$$

定义 1.4 称满足(1.3)的解为该约束的可行解. 当可行解 x 关于 $Ax=b$ 又是基本解,则称它为基本可行解. 对给定的基本可行解 x,若它有 m 个分量大于零,则称 x 为非退化的基本可行解. 否则称为退化的基本可行解. 基本可行解所对应的基称为可行基.

显然,对应于基 B 的基本解为基本可行解的充要条件为

$$B^{-1}b\geqslant 0$$

基本可行解是线性规划中最重要的概念. 现在,我们能够指出,基本可行解与对应的约束(1.3)组成的图形的顶点是等价的. 为此,

定义 1.5 称 R^n 中一个子集 Ω 为凸集,若对$\forall\ x,y\in\Omega$,$\forall\ \theta$: $0<\theta<1$,都有

$$\theta x+(1-\theta)y\in\Omega$$

形象地说,凸集中任意两点间的连线都在该集之中. 容易由定义验证,线性规划之可行集(1.1)为凸集.

定义 1.6 凸集 Ω 中的点 x 称为 Ω 的顶点(或称极点),如果不存在 Ω 中的两个相异的点 y 和 z,使得对于某个 θ: $0<\theta<1$ 有 $x=\theta y+(1-\theta)z$.

定理 1.7(顶点和基本可行解的等价性) x 为 $X=\{x: Ax=b, x\geqslant 0\}$的顶点的充要条件是 x 为(1.3)的一个基本可行解.

证 不妨设 $x=(x_1,\cdots,x_m,0,\cdots,0)^T$ 是(1.3)的一个基本可行解.对应的基为首 m 列.于是

$$x_1a_1+x_2a_2+\cdots+x_ma_m=b$$

其中 $a_1,\cdots,a_m$ 是线性无关的.假定有 $y,z\in X$,使得对某个 θ 有

$$x=\theta y+(1-\theta)z,\ 0<\theta<1$$

由于 y,z 的所有分量均非负,立得 y 和 z 的后 $n-m$ 个分量都为零.此时对应地有

$$y_1a_1+y_2a_2+\cdots+y_ma_m=b$$
$$z_1a_1+z_2a_2+\cdots+z_ma_m=b$$

由 $a_1,\cdots,a_m$ 的线性无关性,只有 $x=y=z$.故 x 为 X 的顶点.

反之,设 x 为 X 的一个顶点.设 x 的非零分量为首 k 个分量.于是

$$x_1a_1+x_2a_2+\cdots+x_ka_k=b \tag{1.4}$$

其中 $x_i>0,i=1,\cdots,k$.我们只须证 $a_1,\cdots,a_k$ 线性无关即可.事实上,若 $a_1,a_2,\cdots,a_k$ 是线性无关的,则在 $k=m$ 时 x 是非退化的基本可行解;在 $k<m$ 时,因 A 之秩为 m,则可从 A 的余下的 $n-k$ 个列中找出 $m-k$ 个列,使之与 $a_1,a_2,\cdots,a_k$ 一起组成一个基,令 x 的相应的 $m-k$ 个分量为零,这样得到一个退化的基本可行解.

今假定 $a_1,a_2,\cdots,a_k$ 是线性相关的,于是存在不全为零的 y_i,$i=1,\cdots,k$,使得

$$y_1a_1+y_2a_2+\cdots+y_ka_k=0 \tag{1.5}$$

记 $y=(y_1,\cdots,y_k,0,\cdots,0)^T\in R^n$,则由(1.4)和(1.5)知有 $A(x\pm\varepsilon y)=b,\forall\ \varepsilon$;又,可选得 $\varepsilon\neq0$ 使

$$x+\varepsilon y\geqslant0\quad 且\quad x-\varepsilon y\geqslant0$$

总之就有 $x\pm\varepsilon y\in X$.于是从

$$x=\frac{1}{2}(x+\varepsilon y)+\frac{1}{2}(x-\varepsilon y)$$

知 x 不是 X 的顶点,这与原设矛盾.这个矛盾说明了 $a_1,a_2,\cdots,a_k$ 是线性无关的.证毕.

例 1.8 写出例 1.2 的约束,并画出关于基本可行解的对应

的图，我们发现它与图 1.1 相重合.

$$\begin{aligned} x_1 + x_2 + x_3 \qquad\qquad &= 3500 \\ x_1 \qquad\quad + x_4 \qquad &= 1500 \\ 5x_1 + 2x_2 \qquad\quad + x_5 &= 10000 \\ x_i \geqslant 0, i = 1,2,3,4,5 & \end{aligned}$$

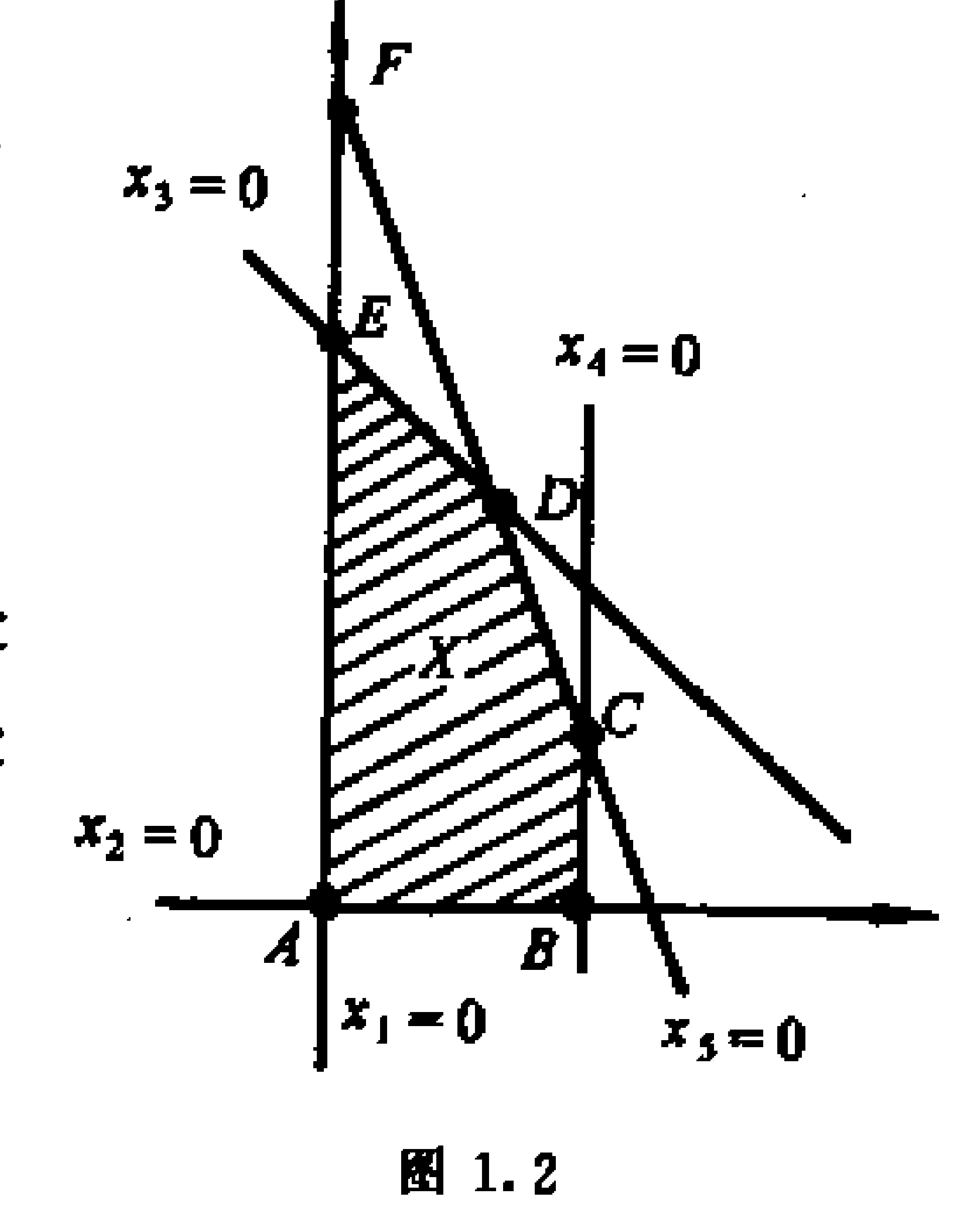


图 1.2

图 1.2 中的每条边对应于一个为零变量，而顶点是某两个变量同时为零所得. 计算得

$A=(0,0,3500,1500,10000)$

$B=(1500,0,2000,0,2500)$

$C=(1500,1250,750,0,0)$

$D=(1000,2500,0,500,0)$ $\qquad$ $E=(0,3500,0,1500,3000)$

它们都是非退化的基本可行解. 注意，还有一些基本解图中未标出，它们是不可行的.

§1.3 线性规划的基本定理

以下，我们只考虑标准型

$$\text{(LP)} \qquad \begin{aligned} &\min c^T x \\ &\text{s.t. } Ax = b \\ &\qquad x \geqslant 0 \end{aligned}$$

定义 1.9 目标函数在约束的假定下达到极小的可行解称最优解. 若这个解还是基本解，则称之为最优基本可行解.

定理 1.10（线性规划的基本定理） 对于线性规划的标准型(LP)，其中 A 为 $m\times n$ 阵，秩为 m，有

(i) 若存在一个可行解，则必存在一个基本可行解；

(ii) 若存在一个最优解，则必存在一个最优基本可行解.

证 证明是构造性的. 假定 x 是可行(或最优)解. 设 x 有 k 个分量大于零，不妨设首 k 个. 于是有

$$x_1a_1 + x_2a_2 + \cdots + x_ka_k = b \tag{1.6}$$

若 $a_1,a_2,\cdots,a_k$ 是线性无关的，则根据(1.4)以下的一段论述知结论已得；故下设 $a_1,a_2,\cdots,a_k$ 线性相关. 于是又有

$$y_1a_1 + y_2a_2 + \cdots + y_ka_k = 0 \tag{1.7}$$

其中 $y_i,i=1,\cdots,k$ 是不全为零的一组数，不妨设其中有一为正.

(i) 设 x 为可行解. 由(1.7)减去(1.6)的 ε 倍得

$$(x_1 - \varepsilon y_1)a_1 + \cdots + (x_k - \varepsilon y_k)a_k = b \tag{1.8}$$

令 $y=(y_1,\cdots,y_k,0,\cdots,0)^T$，由(1.8)推知 $x-\varepsilon y$ 满足方程 $Ax=b$. 为保持可行性，令

$$\varepsilon = \min\{x_i/y_i : i = 1,\cdots,k \text{ 且 } y_i > 0\} \tag{1.9}$$

这样，$x-\varepsilon y$ 中至多只有 $k-1$ 个分量不为零. 必要时重复这一过程，最后得到的可行解其正分量所对应的 A 中的列线性无关.

(ii) 设 x 为最优解. 在(i)的证明的基础上补证目标函数值不变即可. 即证

$$c^T(x - \varepsilon y) = c^Tx \quad \text{或} \quad c^Ty = 0$$

由于 $x_i>0,i=1,\cdots,k$，故对足够小的 $|\varepsilon|\neq 0$ 有 $x\pm\varepsilon y\geqslant 0$，从而可知 $x\pm\varepsilon y$ 对(LP)可行. 再据 x 的最优性得到

$$c^T(x-\varepsilon y)\geqslant c^Tx \Rightarrow \varepsilon c^Ty\leqslant 0$$
$$c^T(x+\varepsilon y)\geqslant c^Tx \Rightarrow \varepsilon c^Ty\geqslant 0$$

从而推及 $c^Ty=0$. 证毕.

基本定理告诉我们：若已知(LP)有最优解，则寻找其最优解只要在基本可行解中寻找即可. 易知基本可行解的个数至多为 C_n^m，这就导致求解线性规划的穷举法. 我们用下面的例子说明.

例 1.11 考虑例 1.2，易知其可行集是非空的有界闭集，故其有最优解. 根据基本定理，将基本可行解分别代入其目标函数中：例 1.8 的 A,B,C,D,E 中以 D 所对应的目标值为最小，故 D 为最优解. 回复到例 1.1，$x=(1000,2500)$是它的最优解.

对于一个给定的线性规划问题，理论上我们已能求出一个最

优解，如果最优解存在的话. 下一个定理本质上给出了全部最优解. 证明可参见文献[24]，但叙述已经修正.

定理 1.12（全部最优解的给出） 设标准型线性规划(LP)的所有最优基本可行解为

$$x^{(1)},x^{(2)},\cdots,x^{(p)}$$

并设(LP)的最优解集有界，则 x 为(LP)的最优解的充要条件是它可以表成上面一组解的凸组合. 即

$$x=\sum_{i=1}^{p}\lambda_i x^{(i)},\quad \lambda_i\geqslant 0\text{ 且}\sum_{i=1}^{p}\lambda_i=1$$

§1.4 实际应用的例子

线性规划可以用来解决科学研究、工程设计、活动安排、军事指挥、经济规划、经营管理等方面提出的大量问题. 此处仅举两例说明.

例 1.13（饮食问题） 一个大的军队的饮食可能会提出这样的问题：假定市场上可以买到几种不同的食品，第 j 种食品的单价为 c_j，$j=1,\cdots,n$. 另外有 m 种营养成分，设每天必须供应第 i 种营养成分至少为 b_i 单位，$i=1,\cdots,m$. 再假定第 j 种食品每单位含有第 i 种营养为 a_{ij} 个单位. 问如何在满足基本营养的条件下购买食品耗资最少？

设 x_j 为第 j 种食品购买的单位数，则

$$\min\ c_1x_1+c_2x_2+\cdots+c_nx_n$$

（购买食品耗资最少）

$$\text{s.t.}\ a_{11}x_1+a_{12}x_2+\cdots+a_{1n}x_n\geqslant b_1$$

$$\cdots\quad\cdots\quad\cdots\quad\cdots$$

$$a_{m1}x_1+a_{m2}x_2+\cdots+a_{mn}x_n\geqslant b_m$$

（应满足基本的营养供给）

$$x_1\geqslant 0,x_2\geqslant 0,\cdots,x_n\geqslant 0$$

（购买食品数均为非负）

用矩阵及向量的形式来写(其中 A,b,c,x 自明),即

$$\min\ c^T x$$
$$\text{s.t.}\ \ Ax \geqslant b$$
$$x \geqslant 0$$

例1.14(运输问题) 某种产品设有 m 个产地,产量分别为 $a_1,a_2,\cdots,a_m$;有销地 n 个,销量分别为 $b_1,b_2,\cdots,b_n$. 设将产品从第 i 个产地运到第 j 个销地的单位运输成本为 c_{ij},对应的运输量为 $x_{ij}(i=1,\cdots,m;j=1,\cdots,n)$. 问如何在满足销量的情况下设计运输方案使运费最低?

$$\min\ \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}x_{ij}$$

(所花的总运费最少)

$$\text{s.t.}\ \ \sum_{j=1}^{n} x_{ij} = a_i,\ i = 1,\cdots,m$$

(各地发出的量等于产量)

$$\sum_{i=1}^{m} x_{ij} = b_j,\ j = 1,\cdots,n$$

(满足各销地的需要)

$$x_{ij} \geqslant 0,\ i = 1,\cdots,m;j = 1,\cdots,n$$

(运出或收到的量均非负)

读者不难看出这是标准型线性规划. 另外,根据模型的具体构造有

$$\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j \tag{1.10}$$

即发出的总量与收到的总量相同.

习　　题

1. 化下述问题为标准型并求解.

$$\begin{aligned} \max\ & x_1+4x_2+x_3 \\ \text{s.t.}\ & 2x_1-2x_2+x_3=4 \\ & x_1 \qquad -x_3=1 \\ & x_2,\ x_3\geqslant 0 \end{aligned}$$

2. 求出

$$\begin{aligned} \min\ & x_1+x_2+5x_3 \\ \text{s.t.}\ & x_1+x_2+x_3=1 \\ & x_1,\ x_2,\ x_3\geqslant 0 \end{aligned}$$

的全部解，并在 R^3 中画出约束集和标出全部解.

3. 将下面的线性规划化为标准型

$$\begin{aligned} \min\ & c^T x \\ \text{s.t.}\ & Ax=b \\ & x\geqslant d \end{aligned}$$

其中 $d\in R^n$，且 $d\neq 0$，即 d 的分量不全为零.

4. 某食品商店生产两种糖果. 第一种每盒获利40分，第二种每盒获利50分. 制作过程主要是混合、烹调、包装三种工作. 下表为每盒糖果在制作过程中所需要的平均时间，以分钟计算.

	混合	烹调	包装
第一种	1分	5分	3分
第二种	2分	4分	1分

混合设备每天至多只能用12机器小时，烹调设备每天至多只能用30机器小时，包装设备每天至多只能用15机器小时. 问如何安排生产获利最大？作图以求其解.

5. 假定标准线性规划(LP)的约束集是非空的，又假定它的目标函数在可行集 X 上是下有界的，则(LP)必有最优解.

6. 如果有 n 个 m 维向量，$n\geqslant m$，而且这 n 个向量中的任意 m 个都线性无关，则称这 n 个向量满足 Haar 条件. 试证约束 $Ax=b, x\geqslant 0$中，若 A 的列向量及 b 这 $n+1$个向量满足 Haar 条件，对应之线性规划没有退化的基本可行解.

7. 假定有 n 个人，n 种工作，不同的人被分派作不同的工作. 设第 i 个人做第 j 种工作须支出 c_{ij}. 问怎样分派这些人的工作使总支出最少？(仅建立模型)

提示：令 x_{ij}表第 i 个人做第 j 种工作. 定义

$$x_{ij}=\begin{cases}1, & \text{若第 } i \text{ 个人做第 } j \text{ 种工作,}\\ 0, & \text{若第 } i \text{ 个人不做第 } j \text{ 种工作.}\end{cases}$$

8. 利用某类钢板下 m 种零件的毛料. 设在一块钢板上已设计出 n 种不同的下料方式;设第 j 种下料方式中可下得第 i 种零件 a_{ij}个. 设第 i 种零件之所需量为 b_i 个. 建立既满足需要又使所用的钢板总数为最少的数学模型.

第二章 单纯形法

基本定理告诉我们，最优解可以在基本可行解里寻找. 但找出全部基本可行解，然后检验谁为最优，当 n 很大同时 m 适量时，运算量极为可观. 事实上，一旦给出某个基本可行解，则我们没有必要再去考虑那些使目标值变大的基本可行解，而是从现有的基本可行解出发去构造使目标值变小的基本可行解. 单纯形法就是基于这种思想的计算方法.

本章的所有讨论都是针对标准型线性规划(LP)的.

§2.1 单纯形法的基本理论

本节，我们从计算的角度来推导单纯形法的基本理论. 请注意各小节的假设条件.

1. 求基本解的迭代法

基本可行解必须是基本解. 要牢固地掌握单纯形法，必须先搞清基本解的迭代求法.

考虑方程组

$$Ax = b \tag{2.1}$$

其中 A 为 $m\times n$ 阶阵，秩为 m，$m<n$. 任意取定 A 的 m 个线性无关列，不妨设为首 m 列. 记 $B=(a_1, a_2, \cdots, a_m)$，则(2.1)可化为等价的方程组

$$B^{-1}Ax = B^{-1}b \tag{2.2}$$

将(2.2)的系数列成表，并以 $y_{ij}, i=1,\cdots,m; j=1,\cdots,n$ 记 $B^{-1}A$ 的第 i,j 元，以 $y_{i0}, i=1,\cdots,m$ 记 $B^{-1}b$ 的第 i 元，则有

$$
\begin{matrix}
1 & 0 & \cdots & 0 & y_{1,m+1} & y_{1,m+2} & \cdots & y_{1n} & y_{10} \\
0 & 1 & \cdots & 0 & y_{2,m+1} & y_{2,m+2} & \cdots & y_{2n} & y_{20} \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & & \vdots & \vdots \\
0 & 0 & \cdots & 1 & y_{m,m+1} & y_{m,m+2} & \cdots & y_{mn} & y_{m0}
\end{matrix} \tag{2.3}
$$

相应的基本解就是$(y_{10},y_{20},\cdots,y_{m0},0,\cdots,0)^T$，其中对应的变量$x_1,x_2,\cdots,x_m$是基本变量，其余的是非基本变量。

设已给出标准化方程组(2.3)。假定要把其中的一个基本变量变为非基本变量，同时将一个非基本变量变为基本变量。比如说要将非基本变量x_k代替某个基本变量x_l，$1\leqslant l\leqslant m$，当且仅当$y_{lk}\neq 0$时才能做到这一点。具体的做法是：先用y_{lk}去除(2.3)的第l行，则其第l,k元变为1，然后再使(2.3)中第k列其余的元化为零。用y'_{ij}，$i=1,\cdots,m$；$j=0,1,\cdots,n$表新方程之系数，则

$$
\begin{cases}
y'_{ij}=y_{ij}-\dfrac{y_{lj}}{y_{lk}}y_{ik}, i=1,\cdots,m, i\neq l; j=0,1,\cdots,n \\
y'_{lj}=\dfrac{y_{lj}}{y_{lk}}, j=0,1,\cdots,n
\end{cases} \tag{2.4}
$$

容易验证，这一变换除了(2.3)中的第l列外，不影响其它基本变量对应的列。而第k列已化为第l个元为1的m维单位列向量。在新的标准方程组的系数阵中能够写出新的基本解，其最后列对应于新的基本变量，但次序已有变动：

$$(y'_{10},\cdots,y'_{l-1,0},0,y'_{l+1,0},\cdots,y'_{m0},0,\cdots,0,y'_{l0},0,\cdots,0)$$

其中y'_{l0}对应于新的基本变量x_k。这样我们实现了一步基本变量的转移。原方程组(2.3)中的元y_{lk}叫做主元。这种运算称取主运算。第k列称为进基向量，第l列称为离基向量。

例2.1 写出例1.2中对应于$Ax=b$的系数阵，即

初 始 表

x_1	x_2	x_3	x_4	x_5	
1	1	1	0	0	3500
1	0	0	1	0	1500
5	(2)	0	0	1	10000

它已是(2.3)的形式。从上表中读出基本解(0,0,3500,1500,

10000). 表中〈·〉所标出的是我们取的主元，即用 x_2 代替 x_5 作为基本变量. 利用(2.4)计算得下表，同时括出下一个主元.

第 一 表

〈−3/2〉	0	1	0	−1/2	−1500
1	0	0	1	0	1500
5/2	1	0	0	1/2	5000

基本解为(0,5000,−1500,1500,0). 用 x_1 代替 x_3，得

第 二 表

1	0	−2/3	0	1/3	1000
0	0	2/3	1	−1/3	500
0	1	5/3	0	1/6	2500

基本解为(1000,2500,0,500,0). 将上述结果与例1.8中的图1.2相对照，知上迭代在图中的行进路线为 $A\to F\to D$.

2. 求基本可行解的迭代法

我们已经知道在研究标准型线性规划时仅需考虑约束

$$\begin{cases}Ax = b\\ x \geqslant 0\end{cases} \tag{2.5}$$

的基本可行解. 前一节的论述说明了取主运算能由一个旧基产生新基，新的基本解仍可由变换了的表(2.3)表述. 一般说来，任意指定一组基 B，对应于 $Ax=b$ 的基本解未必满足非负性. 已经指出，欲对应于基 B 的基本解对(2.5)是可行的即 B 为可行基(见定义1.4)，当且仅当 $B^{-1}b\geqslant 0$. 这一节我们指出，在给定一组基本可行解的情况下，如何通过取主运算来得到一组新的基本可行解，即如何选择进基变量 x_k 和离基变量 x_l.

仍写出标准方程组的增广阵

a_1	a_2	$\cdots$	a_m	a_{m+1}	a_{m+2}	$\cdots$	a_n	b
1	0	$\cdots$	0	$y_{1,m+1}$	$y_{1,m+2}$	$\cdots$	y_{1n}	y_{10}
0	1	$\cdots$	0	$y_{2,m+1}$	$y_{2,m+2}$	$\cdots$	y_{2n}	y_{20}
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
0	0	$\cdots$	1	$y_{m,m+1}$	$y_{m,m+2}$	$\cdots$	y_{mn}	y_{m0}

(2.6)

表的最上行我们现在用 A 的列和 b 来表示，因为现在需要强调列向量之间的关系. 为使新基本解为可行即要求变换后的 b 列各元皆非负，在(2.4)中令 $j=0$得变换后的(2.6)之 b 列为

$$\begin{cases} y'_{i0}=y_{i0}-\dfrac{y_{l0}}{y_{lk}}y_{ik}, i=1,\cdots,m, i\neq l \\ y'_{l0}=\dfrac{y_{l0}}{y_{lk}} \end{cases} \tag{2.7}$$

已知 $y_{i0}\geqslant 0, i=1,\cdots,m$，欲使 $y'_{i0}\geqslant 0, i=1,\cdots,m$. 由(2.7)之第二式可以看出一般应取 $y_{lk}>0$，即对应于 a_k 的列中主元应为一个正数. 其次考虑 $i\neq l$，由(2.7)之第一式可看出

若 $y_{ik}\leqslant 0$，则恒有 $y'_{i0}\geqslant 0$；

若 $y_{ik}>0$，则欲 $y'_{i0}\geqslant 0$，应选 $l\in I_k$：

$$I_k=\{i: y_{i0}/y_{ik}=\min_{i:y_{ik}>0} y_{i0}/y_{ik}, i=1,\cdots,m \text{ 且 } y_{ik}>0\} \tag{2.8}$$

注意，如果集合 I_k 的元多于一个(这种情况有的书中称之为平局或平结)，此时 l 的选择不唯一. 变换(2.4)必然会导致退化的基本可行解，这表现在在变换后的(2.6)中对应于 b 列的元有为零者.

我们用例来说明上面的方法.

例2.2 重新写出例2.1中的初始表

初 始 表

a_1	a_2	a_3	a_4	a_5	b
1	(1)	1	0	0	3500
1	0	0	1	0	1500
5	2	0	0	1	10000

现行的基本可行解为(0,0,3500,1500,10000). 今选 a_2进基，计算有关比值3500/1,10000/2，选取非负数中的最小者，其对应于第一行，故取1为主元，表中已括出. 用 a_2代替 a_3进基，新表为

第 一 表

1	1	1	0	0	3500
1	0	0	1	0	1500
(3)	0	−2	0	1	3000

基本可行解为(0,3500,0,1500,3000). 再选 a_1 进基，表中已括出主元. 进一步得

第 二 表

0	1	5/3	0	−1/3	2500
0	0	2/3	1	−1/3	500
1	0	−2/3	0	1/3	1000

基本可行解为(1000,2500,0,500,0). 读者可以看出此例对应于例1.8中图1.2的进程为 $A \to E \to D$.

在表(2.6)中，若某 a_k 列中无正的元，我们自然不选 a_k 进基，但这种情况是有特殊意义的：它表示(2.6)所对应的方程组(2.5)的可行解集无界. 兹证如下：

依(2.6)的记号，有

$$y_{10}a_1 + y_{20}a_2 + \cdots + y_{m0}a_m = b \tag{2.9}$$

类似地，在(2.6)中 a_k 列与 b 列有相似的地位，故亦有

$$y_{1k}a_1 + y_{2k}a_2 + \cdots + y_{mk}a_m = a_k \tag{2.10}$$

将(2.9)减去(2.10)的 ε 倍，整理得

$$(y_{10} - \varepsilon y_{1k})a_1 + \cdots + (y_{m0} - \varepsilon y_{mk})a_m + \varepsilon a_k = b \tag{2.11}$$

因 a_k 列中各元皆非正并且诸 y_{i0} 皆非负，上式表明对 $\forall\ \varepsilon \geqslant 0$,

$$(y_{10} - \varepsilon y_{1k} \cdots, y_{m0} - \varepsilon y_{mk}, 0, \cdots, 0, \varepsilon, 0, \cdots, 0) \tag{2.12}$$

其中分量 ε 对应于 x_k 的位置，都是(2.5)的解. 从而证明了(2.5)的解集无界.

总之，设已有了一个基本可行解，则可通过取主运算得到新的基本可行解；或能知道问题的可行解集是无界的，且得到了形如(2.12)的可行解.

3. 基本可行解的改进

上节中选择的进基向量 a_k 仍带有一定的随意性. 本节，我们指出如何按一定的规则确定进基向量 a_k 来得到一个新的基本可行解，同时使目标函数值下降(至少不增)；或者确定目标函数在可行集上是无下界的(当然，对应的线性规划就没有最优解).

假定已知基本可行解为

$$x=(y_{10},y_{20},\cdots,y_{m0},0,\cdots,0)^T$$

对应的增广阵应是

	a_1	a_2	$\cdots$	a_m	a_{m+1}	a_{m+2}	$\cdots$	a_n	b
c_1	1	0	$\cdots$	0	$y_{1,m+1}$	$y_{1,m+2}$	$\cdots$	y_{1n}	y_{10}
c_2	0	1	$\cdots$	0	$y_{2,m+1}$	$y_{2,m+2}$	$\cdots$	y_{2n}	y_{20}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
c_m	0	0	$\cdots$	1	$y_{m,m+1}$	$y_{m,m+2}$	$\cdots$	y_{mn}	y_{m0}

(2.13)

对应于任一可行解 x，目标函数值为

$$s=c^Tx=c_1x_1+c_2x_2+\cdots+c_nx_n \tag{2.14}$$

对应于现行的基本可行解的目标函数值为

$$s_0=c_1y_{10}+c_2y_{20}+\cdots+c_my_{m0}$$

即表(2.13)中 b 列各元与附加的 c 列各对应元乘积之和. 另一方面，对于任一可行解 $x=(x_1,x_2,\cdots,x_n)^T$，利用(2.13)容易写出：

$$\begin{cases} x_1=y_{10}-\sum\limits_{j=m+1}^{n}y_{1j}x_j \\ x_2=y_{20}-\sum\limits_{j=m+1}^{n}y_{2j}x_j \\ \cdots\qquad\cdots \\ x_m=y_{m0}-\sum\limits_{j=m+1}^{n}y_{mj}x_j \end{cases} \tag{2.15}$$

将(2.15)代入(2.14)并利用 s_0 的表达式，我们得到任意目标函数值 s 与原目标函数值 s_0 之间的一种关系：

$$s=s_0+(c_{m+1}-z_{m+1})x_{m+1}+\cdots+(c_n-z_n)x_n \tag{2.16}$$

其中

$$z_j=\sum_{i=1}^{m}c_iy_{ij},\ j=m+1,\cdots,n$$

今注意到上式中之 z_j 是用(2.13)中的 c 列各元与第 j 列的各对应元的乘积取和而得，因此可以定义

$$z_j=c_j,\ j=1,\cdots,m$$

从而(2.16)亦可表为

$$s = s_0 + \sum_{j=1}^{n}(c_j - z_j)x_j \tag{2.17}$$

并且对于给定的一组基本可行解及对应的表(2.13)，s_0，z_j，$j=1$，$\cdots$，n 都已确定. 它们与变量 x 无关.

我们将已有的讨论总结为以下诸定理.

定理2.3(判别定理)　若对一个基本可行解，有

$$c_j - z_j \geqslant 0,\ j = 1,2,\cdots,n \tag{2.18}$$

则此基本可行解为最优解.

证　让 x 在可行集内任意变动，由 x 的非负性以及条件(2.18)结合(2.17)知 $s \geqslant s_0$，故现行的基本可行解为最优解.

判别定理的一个逆命题在§2.5中给出.

定理2.4(基本可行解的改进)　如果对于某一基本可行解，有某 k 使 $c_k - z_k < 0$. 若对应于(2.13)第 k 列各元中有大于零者，则将 a_k 作为进基向量所得的新基本可行解的目标函数值同原目标函数值相比不增.

证　设新基本可行解为

$$(x_1, x_2, \cdots, x_m, 0, \cdots, 0, x_k, 0, \cdots, 0)^T$$

对应于(2.13)，其中 $x_1, x_2, \cdots, x_m$ 中离基向量对应的元为零. 注意到 $x_k \geqslant 0$，由(2.16)得

$$s = s_0 + (c_k - z_k)x_k \leqslant s_0$$

证毕.

若在上式中有 $x_k > 0$，则得 $s < s_0$. 于是我们给出

非退化性假定　$Ax=b$，$x \geqslant 0$的任一基本可行解都有 m 个严格为正的分量.

推论2.5　在非退化性假定下，定理2.4结论中之“不增”可改进为“严格下降”，并且定理2.3的逆也成立.

定理2.6　如果对某一基本可行解，有某 k 使 $c_k - z_k < 0$. 若对应于(2.13)的第 k 列的各元皆非正数，则问题的目标函数在可行集上是无下界的.

证 考虑到由(2.12)所确定的一组可行解

$$(y_{10}-\varepsilon y_{1k},\cdots,y_{m0}-\varepsilon y_{mk},0,\cdots,0,\varepsilon,0,\cdots,0)^T$$

其中 $x_k=\varepsilon$，此时对应于(2.16)就有

$$s=s_0+(c_k-z_k)\varepsilon$$

由条件及(2.12)下面的一段话知 ε 可任意大. 再考虑到条件 $c_k-z_k<0$ 知结论为真. 证毕.

从以上的论证可以看到 $c_j-z_j,j=1,\cdots,n$ 有特殊的意义. 记

$$r_j=c_j-z_j,j=1,\cdots,n$$

并称之为判别数. 应该注意到，对于一般的标准型线性规划(LP)，判别数由 c，A 和基 B 一意决定，并且对应于基本变量的判别数皆为零.

称判别数皆非负时所对应的可行基为最优基.

定理2.3说明了一个现行的基本可行解何时为最优. 定理2.6说明了目标函数在可行集上何时无下界. 定理2.4说明了一个现行的基本可行解何时可以得到改进. 线性规划的单纯形法就是以这三个定理为基础的. 因此它们是单纯形法的基本理论.

§2.2 单纯形法

在前几节中，我们已经建立了推导单纯形法的理论和技巧. 这里，我们只要将它们联系起来便可得到线性规划的单纯形方法. 我们先建立单纯形表，给出手算的例子，然后再给出单纯形法一般算法的描述.

单纯形表的初始形式如下，仍设前 m 列为基.

	a_1	a_2	$\cdots$	a_m	a_{m+1}	a_{m+2}	$\cdots$	a_n	b
c_1	1	0	$\cdots$	0	$y_{1,m+1}$	$y_{1,m+2}$	$\cdots$	y_{1n}	y_{10}
c_2	0	1	$\cdots$	0	$y_{2,m+1}$	$y_{2,m+2}$	$\cdots$	y_{2n}	y_{20}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
c_m	0	0	$\cdots$	1	$y_{m,m+1}$	$y_{m,m+2}$	$\cdots$	y_{mn}	y_{m0}
	c_1	c_2	$\cdots$	c_m	c_{m+1}	c_{m+2}	$\cdots$	c_n	0
	0	0	$\cdots$	0	$c_{m+1}-z_{m+1}$	$c_{m+2}-z_{m+2}$	$\cdots$	c_n-z_n	$-s_0$

(2.19)

与此表相应的基本可行解为$(y_{10},y_{20},\cdots,y_{m0},0,\cdots,0)^T$. 表中最末行有 n 个判别数. 右下角为现行的目标值加负号. 对照(2.16)及 s_0 的表达式,最末行可以这样得到:将最左列的 c_i 与各对应行相乘,再用最后第二行依次减去相乘后的各行即得. 其实,我们可以去掉表(2.19)中之最左列,只要将最后第二行依次减去其上各行的适当的倍数,且使对应于基本变量的判别数为零则可. 这样得

a_1	a_2	$\cdots$	a_m	a_{m+1}	a_{m+2}	$\cdots$	a_n	b
1	0	$\cdots$	0	$y_{1,m+1}$	$y_{1,m+2}$	$\cdots$	y_{1n}	y_{10}
0	1	$\cdots$	0	$y_{2,m+1}$	$y_{2,m+2}$	$\cdots$	y_{2n}	y_{20}
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
0	0	$\cdots$	1	$y_{m,m+1}$	$y_{m,m+2}$	$\cdots$	y_{mn}	y_{m0}
0	0	$\cdots$	0	r_{m+1}	r_{m+2}	$\cdots$	r_n	$-s_0$

(2.20)

对照上表中最末行判别数,如果 $r_j \geqslant 0, j=1,\cdots,n$,则现行的解为最优,最优值为 s_0. 否则目标值可以改进,或确定目标函数无下界.

设由(2.20)经过了一次取主运算,第 k 列进了基同时第 l 列离了基,$1\leqslant l\leqslant m$. 于是可从新建立的(2.19)表求得一组新判别数,从而得到新的(2.20)表. 用旧(2.20)中的符号加一撇表示新(2.20)表中对应的符号. 类似于(2.16)中 z_j 之定义式,有

$$r_j = c_j - \sum_{i=1}^{m} c_i y_{ij}, j = 0,1,\cdots,n \tag{2.21}$$

$$r'_j = c_j - \left(\sum_{\substack{i=1 \\ i\neq l}}^{m} c_i y'_{ij} + c_k y'_{lj} \right), j = 0,1,\cdots,n \tag{2.22}$$

注意,在上两式中我们取 $j=0$ 且令 $c_0=0$,得 $r_0=-s_0, r'_0=-s'_0$. 利用(2.4)由(2.22)计算得

$$\begin{aligned} r'_j &= c_j - \left[\sum_{\substack{i=1 \\ i\neq l}}^{m} c_i \left(y_{ij} - \frac{y_{lj}}{y_{lk}} y_{ik} \right) + c_k \frac{y_{lj}}{y_{lk}} \right] \\ &= c_j - \left[\sum_{i=1}^{m} c_i \left(y_{ij} - \frac{y_{lj}}{y_{lk}} y_{ik} \right) + c_k \frac{y_{lj}}{y_{lk}} \right] \end{aligned}$$

$$= \left(c_j - \sum_{i=1}^{m} c_i y_{ij}\right) - \frac{y_{lj}}{y_{lk}}\left(c_k - \sum_{i=1}^{m} c_i y_{ik}\right)$$

$$= r_j - \frac{y_{lj}}{y_{lk}} r_k$$

于是就有

$$r_j' = r_j - y_{lj}' r_k,\ j = 0, 1, \cdots, n \tag{2.23}$$

与(2.4)相比便知(2.23)提供了由 r_j 计算 r_j' 的简便方法. 注意到由(2.23)直接得到 $r_k'=0$，手算时的实施法则是，将旧(2.20)表的最末行减去新(2.20)表的第 l 行的适当的倍数，使得对应于新基的元为零. 这样得到新(2.20)表的最末行.

例2.7 重写例2.2的初始表，并将例1.2中目标函数之系数作为最末行添入，同时右下角置零，得

初 始 表

a_1	a_2	a_3	a_4	a_5	b
1	⟨1⟩	1	0	0	3500
⟨1⟩	0	0	1	0	1500
5	2	0	0	1	10000
−5	−3	0	0	0	0

注意，此问题目标函数中 $c_3=c_4=c_5=0$，上表已为(2.20)的形式. 表中最末行判别数有两个为负且对应之 a_1，a_2列中均有正元，将 a_1，a_2中的任一个作为进基向量都会得到改进的解，对应的主元在表中已括出. 例如，我们选 a_2进基，作一次取主运算；同时将最末行作相应的变换，使得 a_2，a_4，a_5下的三个判别数为零，这样得

第 一 表

1	1	1	0	0	3500
1	0	0	1	0	1500
⟨3⟩	0	−2	0	1	3000
−2	0	3	0	0	10500

选 a_1进基，括出主元，得

第 二 表

0	1	5/3	0	−1/3	2500
0	0	2/3	1	−1/3	500
1	0	−2/3	0	1/3	1000
0	0	2/3	0	2/3	12500

最末行之判别数已无为负者，故由第二表得最优解(1000,2500,0,500,0)，最优值为−12500. 回复到例1.1，得最优解(1000,2500)，最优值为12500.

读者也可以在初始表中选a_1进基，然后完成余下相应的计算. 将每表所得之基本可行解与例1.8中的图1.2相对照，看看行进的线路又是如何.

以下给出一般情况的单纯形算法. 再写出标准型

$$\begin{aligned}&\min\ c^Tx\\ &\text{s.t.}\ \ x_1a_1+x_2a_2+\cdots+x_na_n=b\\ &\qquad x\geqslant 0\end{aligned}\tag{2.24}$$

设线性规划(2.24)已有一个基本可行解，对应的基为$a_{j_1},a_{j_2},\cdots,a_{j_m}$. 记

$$B=(a_{j_1},a_{j_2},\cdots,a_{j_m})$$

求出

$$(y_{10},y_{20},\cdots,y_{m0})^T=B^{-1}b$$

$$(y_{1j},y_{2j},\cdots,y_{mj})^T=B^{-1}a_j,\ j=1,\cdots,n$$

注意，$B^{-1}a_{j_i}=e_i,i=1,\cdots,m$. 这里$e_i$表第$i$个元为1的单位$m$维向量. 于是对应的单纯形表为

(2.25)

		a_1	a_2	$\cdots$	a_n	b
j_1	c_{j_1}	y_{11}	y_{12}	$\cdots$	y_{1n}	y_{10}
j_2	c_{j_2}	y_{21}	y_{22}	$\cdots$	y_{2n}	y_{20}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
j_m	c_{j_m}	y_{m1}	y_{m2}	$\cdots$	y_{mn}	y_{m0}
		c_1-z_1	c_2-z_2	$\cdots$	c_n-z_n	$-z_0$

这个表中，当 $j_i=i,i=1,\cdots,m$ 时，就化为(2.20)表.上例中的初始表、第一表及第二表都是这个形式.表中对应于基的列组成一个 m 阶单位阵(尽管次序要经过适当调整，不以下标的大小为序，而以子列标号 j_i 的下标 i 的大小为序).同时末行判别数中对应于基的元为零.

单纯形算法　对于标准型线性规划(2.24)，设已有一个基本可行解及对应的表(2.25)中的 $y_{ij},i=1,\cdots,m;j=0,1,\cdots,n$.

第一步　定主元列 k：

$$z_j=\sum_{i=1}^{m}c_{j_i}y_{ij},j=1,\cdots,n$$

取 k：

$$r_k=c_k-z_k=\min_{1\leqslant j\leqslant n}\{c_j-z_j\}$$

若 $r_k\geqslant 0$，则得最优解 x，其中 $x_{j_i}=y_{i0},i=1,\cdots,m,x$ 的其余分量为零.最优值为

$$s=\sum_{i=1}^{m}c_{j_i}y_{i0}$$

算法结束.否则，

第二步　定主元行 l：

若 $y_{ik}\leqslant 0,i=1,\cdots,m$，则目标函数在可行集上无下界，算法结束；否则，取 l：

$$\frac{y_{l0}}{y_{lk}}=\min_{y_{ik}>0}\left\{\frac{y_{i0}}{y_{ik}}\right\}\tag{2.26}$$

第三步　取主运算和变动标号内容

$$\begin{cases}y'_{lj}=\dfrac{y_{lj}}{y_{lk}},j=0,1,\cdots,n\\ y'_{ij}=y_{ij}-y_{ik}\cdot y'_{lj}\end{cases}\tag{2.27}$$

$$i=1,\cdots,m,i\neq l;j=0,1,\cdots,n$$

$$j_l\Leftarrow k$$

$$c_{j_l}\Leftarrow c_k$$

转回第一步.

在非退化性假定成立的情况下，上算法能在有限步内结束. 这是因为非退化性假定保证了目标函数值在算法中严格下降(推论2.5)，从而算法中产生的基各不相同，而基的个数有限. 于是，或目标值无限(在第二步上结束)，或有限步内得解(在第一步上结束).

在退化的情况下可能会出现基的循环，就是在迭代中总是那么几组基轮番出现. 这是因为目标值不能保证严格下降. 实际问题中还没有人碰到这种循环现象，但有人构造出这种例子. 普遍认为，可仍用以上算法对一般问题求解. 碰到循环情况时再用补救办法，详情参见 § 2.5.

§ 2.3　初始基本可行解的寻求

以上的讨论总是在给定了第一个基本可行解的假定下进行的. 但初始基本可行解如何寻求?利用一个辅助的线性规划与单纯形法相结合，会产生满意的效果.

对于线性规划的初始基本可行解，有时可以直接得到. 例如在以

$$\begin{cases} Ax \leqslant b \\ x \geqslant 0 \end{cases}$$

为约束的问题中，其中 $b \geqslant 0$，对应于这个问题的初始基本可行解可以直接由松弛变量提供. 前节的例已经作了很好的说明. 现在我们研究一般标准型的约束

$$\begin{cases} Ax = b \\ x \geqslant 0 \end{cases} \tag{2.28}$$

其中 $b \geqslant 0$. 这里，我们对 A 的秩不作假定. 为了寻求(2.28)的一个基本可行解，我们研究人造的极小问题：

$$\begin{aligned} \min s &= \sum_{i=1}^{m} y_i \\ \text{s.t. } Ax + Iy &= b \\ x \geqslant 0&, y \geqslant 0 \end{aligned} \tag{2.29}$$

其中 $y=(y_1,y_2,\cdots,y_m)^T$ 是人工向量. 注意,(2.29)已为我们提供了一组基本可行解 $x=0,y=b$. 并且,(2.29)是一个标准的具有 $n+m$ 个变元的线性规划,零是目标函数的一个下界. 因此,(2.29)必有解(上章习题5);于是我们可用单纯形法解之. 有下述可能性:

(i) 若 $\min s>0$,则(2.28)无可行解. 这是因为若(2.28)有可行解 $\bar{x}$,则 $x=\bar{x},y=0$ 便是(2.29)的可行解. 从而对应地有 $s=0$,得矛盾.

(ii) 若 $\min s=0$. 此时有两种可能性.

(a) 若此时基本变量全为 x 变量,即通过单纯形法中基本变量的转移,关于 y 的变量已全部离基. 在(2.29)现行的单纯形表中去掉附加的 y 的 m 列,这样便得到(2.28)的一组基本可行解.

(b)若此时基本变量中含有 y 变量,比如说含有 y_r,那么现行的单纯形表的第 r 行中,对应于 y_r 的系数是1,对应于 b 列的系数是零,其余的对应于 x,y 作为基本变量的系数(共 $m-1$ 个)亦为零. 写出对应的方程应是

$$y_r+\sum_{t\in T}y_{rt}y_t+\sum_{j\in J}y_{rj}x_j=0$$

其中 y_r 为基变量;$y_t,t\in T,x_j,j\in J$ 均为非基变量. 这时又分两种情况:

若 $y_{rj}=0,\forall\ j\in J$,这说明经过行的变换,(2.28)中之 $Ax=b$ 的第 r 个方程的系数及右端全为零,则该第 r 个方程可以去掉. 现行的单纯形表中(对应于(2.29)的)第 r 行即可去之.

若某 $y_{rs}\neq 0,s\in J$,可选择 x_s 进基,取代 y_r. 两种情况的任一种都导致 y_r 消失,从而在有限步内必导致情况(a).

总之,我们使用人工变量方法于(2.28)可以得到下列结果.

(i) (2.28)无可行解;

(ii) (2.28)中的方程组 $Ax=b$ 中有多余的方程,以及决定那些方程可以去掉;

(iii) 得到(2.28)(或化约后的等价的(2.28))的一组基本可行解.

例2.8 研究问题

$$\min\ 4x_1 + x_2 + x_3$$

$$\text{s.t.}\ \ 2x_1 + x_2 + 2x_3 = 4$$

$$3x_1 + 3x_2 + x_3 = 3$$

$$x_1, x_2, x_3 \geqslant 0$$

下面的解法叫做两阶段单纯形法.

阶段 I 引入人工变量 x_4, x_5，求上问题的初始基本可行解. 辅助规划为

$$\min\ x_4 + x_5$$

$$\text{s.t.}\ \ 2x_1 + x_2 + 2x_3 + x_4 \qquad = 4$$

$$3x_1 + 3x_2 + x_3 \qquad + x_5 = 3$$

$$x_1, x_2, x_3, x_4, x_5 \geqslant 0$$

初始表应是

初 始 表

x_1	x_2	x_3	x_4	x_5	b
2	1	2	1	0	4
3	3	1	0	1	3
0	0	0	1	1	0

表中最末行是目标函数之系数，b 下记为零. 为运用单纯形法，变换最末行，使对应于基本变量 x_4, x_5 下面的元为零. 这样得到

第 一 表

2	1	2	1	0	4
(3)	3	1	0	1	3
−5	−4	−3	0	0	−7

在对应于末行具有最负（b 列除外）分量的列中取主元，通过一次取主变换及相应的末行变换，得

第 二 表

0	−1	⟨4/3⟩	1	−2/3	2
1	1	1/3	0	1/3	1
0	2	−4/3	0	5/3	−2

在第二表中只有一种取主方式. 括出主元，经过变换得

第 三 表

0	−3/4	1	3/4	−1/2	3/2
1	5/4	0	−1/4	1/2	1/2
0	0	0	1	1	0

两个人工变量已被移出基外，目标函数值化为零同时得到原始问题的一个基本可行解 $x_1=1/2, x_2=0, x_3=3/2$.

阶段 II 运用从阶段 I 中导出的基本可行解，将原始问题用单纯形法极小化. 即，删去人工变量，取出第三表中以虚线勾出的数据，再添上原目标函数中的系数及零作为最末行. 如此得初始表：

初 始 表

x_1	x_2	x_3	b
0	−3/4	1	3/2
1	5/4	0	1/2
4	1	1	0

变换最末行，使对应于基列的分量为零，我们得到

第 一 表

0	−3/4	1	3/2
1	⟨5/4⟩	0	1/2
0	−13/4	0	−7/2

作取主运算，得

第 二 表

3/5	0	1	9/5
4/5	1	0	2/5
13/5	0	0	−11/5

判别数已均非负．故得最优解(0,2/5,9/5)．最优值为11/5.

§2.4 修正单纯形法

至此，读者或许明白，单纯形法中每次的取主运算只用到极少量的列和行，其它的列和行并不明显地用到．特别是当算法在少量几步便结束时(实际经验表明，在经过大约 m 次或$3m/2$次取主运算就可期望这个方法能得到最优解)．对其它的列和行同时进行运算，在某种意义上讲是一种浪费．修正单纯形法的目的就在于避免这种不必要的计算.

1. 单纯形法的矩阵形式

我们先导出矩阵形式的单纯形表．它可以帮助对单纯形法加深理解，同时可以给出单纯形法的简单描述.

考虑标准型线性规划

$$\begin{aligned} \min\ & c^T x \\ \text{s.t.}\ & Ax = b \\ & x \geqslant 0 \end{aligned} \tag{2.30}$$

假定 A 的首 m 列线性无关，记 $B=(a_1,a_2,\cdots,a_m)$，将矩阵 A 分块为

$$A = (B,D)$$

对应地有

$$x = \begin{pmatrix} x_B \\ x_D \end{pmatrix},\quad c = \begin{pmatrix} c_B \\ c_D \end{pmatrix}$$

标准型(2.30)即化为

$$
\begin{aligned}
&\min\ c_B^T x_B + c_D^T x_D \\
&\text{s.t.}\ \ Bx_B + Dx_D = b \\
&\qquad x_B \geqslant 0,\ x_D \geqslant 0
\end{aligned}
\tag{2.31}
$$

由(2.31)的等式立得

$$
x_B = B^{-1}b - B^{-1}Dx_D \tag{2.32}
$$

令 $x_D=0$，由上式得一组基本解. 这里假定解是可行的，即基本可行解为$\begin{pmatrix} x_B \\ 0 \end{pmatrix}$，其中 $x_B=B^{-1}b\geqslant 0$. 对应的目标函数值为 $c_B^T B^{-1}b$.

现在，让 x 在可行集内任意变动，则利用(2.32)，目标函数 $s=c^T x$ 的表达式为

$$
\begin{aligned}
s &= c_B^T x_B + c_D^T x_D \\
&= c_B^T B^{-1}b + (c_D^T - c_B^T B^{-1}D)x_D
\end{aligned}
\tag{2.33}
$$

其中 $n-m$ 维行向量 $c_D^T-c_B^T B^{-1}D$ 就是 $n-m$ 个判别数，它们对应于非基变量. (2.30)和(2.31)有关的系数阵为

$$
\begin{pmatrix} A & b \\ c^T & 0 \end{pmatrix} = \begin{bmatrix} B & D & b \\ c_B^T & c_D^T & 0 \end{bmatrix}
$$

对应的单纯形表为

$$
T_B = \begin{bmatrix} I & B^{-1}D & B^{-1}b \\ 0 & c_D^T - c_B^T B^{-1}D & -c_B^T B^{-1}b \end{bmatrix} \tag{2.34}
$$

它就是(2.20)的矩阵表示. 注意，我们以 T_B 记之，表明(2.34)由 B 一意决定，并且

$$
T_B = \begin{bmatrix} B^{-1} & 0 \\ -c_B^T B^{-1} & 1 \end{bmatrix} \begin{bmatrix} B & D & b \\ c_B^T & c_D^T & 0 \end{bmatrix}
$$

等式右端之第一阵是对应的变换阵.

今假定基 B 由 A 的任 m 个线性无关列所组成. 记

$$
B = (a_{j_1}, a_{j_2}, \cdots, a_{j_m})
$$

对应地记

$$
\begin{aligned}
x_B &= (x_{j_1}, x_{j_2}, \cdots, x_{j_m})^T \\
c_B &= (c_{j_1}, c_{j_2}, \cdots, c_{j_m})^T
\end{aligned}
$$

仿上，我们能够容易地导出一般形式的单纯形表为

$$T_B = \begin{pmatrix} B^{-1}A & B^{-1}b \\ c^T - c_B^T B^{-1}A & -c_B^T B^{-1}b \end{pmatrix} \tag{2.35}$$

它是(2.25)的矩阵形式. 注意，

$$T_B = \begin{pmatrix} B^{-1} & 0 \\ -c_B^T B^{-1} & 1 \end{pmatrix} \begin{pmatrix} A & b \\ c^T & 0 \end{pmatrix}$$

显然，(2.34)是(2.35)的特殊形式.

2. 修正单纯形法

现在，我从单纯形表(2.35)结合(2.34)出发，分析§2.2所给出的单纯形算法. 我们发现首先感兴趣的是由判别数组成的向量 $c_D^T - c_B^T B^{-1}D$. 若 $c_D^T - c_B^T B^{-1}D \geqslant 0$，则已得最优解. 否则，我们找出 $c_D^T - c_B^T B^{-1}D$ 中的最负分量，比如说其下标为 k，它对应于进基向量 a_k. 于是我们再寻找现行的 b 列与 a_k 列的对应分量的比的非负数中之最小者，设为对应于 a_k 列的第 l 元，则选 y_{lk} 为主元，或者确定目标函数无下界. 每次迭代在取主运算之前所涉及的是两列(a_k 列和 b 列)与末行中之 $c_D^T - c_B^T B^{-1}D$. 修正单纯形法的思想是这样的：需用的数据用合适的方法先算出来，(暂时)不用的数据则存放原始数据. 在计算机中实现这种算法一般可节省存储，减少计算量. 现在我们来描述算法且设标准型线性规划(2.30)为给定.

修正单纯形算法　给出初始基的逆 B^{-1} 及现行的 $y_0 = B^{-1}b$ 之后，执行以下步骤：

第一步　计算 $c_D^T - c_B^T B^{-1}D$. 若 $c_D^T - c_B^T B^{-1}D \geqslant 0$，算法结束，得最优解 x 为

$$x_B = y_0,\ x_D = 0$$

第二步　根据 $c_D^T - c_B^T B^{-1}D$ 中的最负数来确定进基向量，设标号为 k. 计算 $y_k = B^{-1}a_k$，这里用 a_k 表示初始表的第 k 列，用 y_k 表示现行表的第 k 列.

第三步　计算比 $\alpha_i = y_{i0}/y_{ik}$，$i=1,\cdots,m$，若 α_l 是诸 α_i 中非负

数的最小者，则主元为 y_{lk}；或确定目标函数是无下界的.

第四步　修改 B^{-1} 和现行的 $y_0=B^{-1}b$，返回第一步.

以下作几点说明：

(i) 引进松弛变量或对标准型使用人工变量方法求得一初始表如(2.35). 因此，初始的 $B^{-1}=I$.

(ii) 对现行的基作取主变换的变换阵由(2.4)式得到：

$$E_{lk}=\begin{pmatrix} 1 & 0 & \cdots & 0 & -y_{1k}/y_{lk} & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & -y_{2k}/y_{lk} & 0 & \cdots & 0 \\ & \cdots & & & \cdots & & \cdots & \\ 0 & 0 & \cdots & 1 & -y_{l-1,k}/y_{lk} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 1/y_{lk} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & -y_{l+1,k}/y_{lk} & 1 & \cdots & 0 \\ & \cdots & & & \cdots & & \cdots & \\ 0 & 0 & \cdots & 0 & -y_{mk}/y_{lk} & 0 & \cdots & 1 \end{pmatrix} \tag{2.36}$$

若以 E_r 记第 r 次的变换阵，则现行的

$$B^{-1}=E_r\cdot E_{r-1}\cdots E_2\cdot E_1$$

在电子计算机中存储这些阵很简单，只要存储取主变换序号 r、对应之标号 l 以及对应的(2.36)中第 l 列的数据. 并且算法中有关的计算也很简单：

$$B^{-1}b=E_r(E_{r-1}\cdots(E_2(E_1b))\cdots)$$

$$B^{-1}a_k=E_r(E_{r-1}\cdots(E_2(E_1a_k))\cdots)$$

若记

$$\lambda^T=c_B^TB^{-1}=(\cdots((c_B^TE_r)E_{r-1})\cdots E_2)E_1$$

则

$$c_D^T-c_B^TB^{-1}D=c_D^T-\lambda^TD$$

(iii) 在引进人工变量求初始基本可行解时，自然也可以使用修正单纯形法.

以上若干说明的实施，有的书中称之为逆乘积形式的单纯形法.

§2.5 摄动理论及避免循环

前面我们已经给出了求解线性规划的单纯形法，并且指出对一般的实际问题总能在有限步内求解. 然而，出现循环的可能依然存在. E. Beale 在1955年给出了如下的例.

例2.9

$$\min -\frac{3}{4}x_1+150x_2-\frac{1}{50}x_3+6x_4$$

$$\text{s.t.}\quad \begin{aligned} \frac{1}{4}x_1-60x_2-\frac{1}{25}x_3+9x_4+x_5 &= 0\\ \frac{1}{2}x_1-90x_2-\frac{1}{50}x_3+3x_4+x_6 &= 0\\ x_3+x_7 &= 1\\ x_i\geqslant 0,\ i=1,2,3,4,5,6,7 & \end{aligned}$$

易知，

$$\left(0,\ 0,\ 1,\ 0,\ \frac{1}{25},\ \frac{1}{50},\ 0\right) \tag{2.37}$$

是这问题的一个可行解，对应的目标函数值为 $-\frac{1}{50}$. 今写出它的初始表

初 始 表

a_1	a_2	a_3	a_4	a_5	a_6	a_7	b
$\langle\frac{1}{4}\rangle$	-60	$-\frac{1}{25}$	9	1	0	0	0
$\frac{1}{2}$	-90	$-\frac{1}{50}$	3	0	1	0	0
0	0	1	0	0	0	1	1
$-\frac{3}{4}$	150	$-\frac{1}{50}$	6	0	0	0	0

以下我们给出迭代表格但略去计算过程，主元都取在判别数最小者所对应的列中.

第 一 表

1	−240	$-\frac{4}{25}$	36	4	0	0	0
0	(30)	$\frac{3}{50}$	−15	−2	1	0	0
0	0	1	0	0	0	1	1
0	−30	$-\frac{7}{50}$	33	3	0	0	0

第 二 表

1	0	$(\frac{8}{25})$	−84	−12	8	0	0
0	1	$\frac{1}{500}$	$-\frac{1}{2}$	$-\frac{1}{15}$	$\frac{1}{30}$	0	0
0	0	1	0	0	0	1	1
0	0	$-\frac{2}{25}$	18	1	1	0	0

第 三 表

$\frac{25}{8}$	0	1	$-\frac{525}{2}$	$-\frac{75}{2}$	25	0	0
$-\frac{1}{160}$	1	0	$(\frac{1}{40})$	$\frac{1}{120}$	$-\frac{1}{60}$	0	0
$-\frac{25}{8}$	0	0	$\frac{525}{2}$	$\frac{75}{2}$	−25	1	1
$\frac{1}{4}$	0	0	−3	−2	3	0	0

第 四 表

$-\frac{125}{2}$	10500	1	0	(50)	−150	0	0
$-\frac{1}{4}$	40	0	1	$\frac{1}{30}$	$-\frac{2}{3}$	0	0
$\frac{125}{2}$	−10500	0	0	−50	150	1	1
$-\frac{1}{2}$	120	0	0	−1	1	0	0

第 五 表

$-\frac{5}{4}$	210	$\frac{1}{50}$	0	1	-3	0	0
$\frac{1}{6}$	-30	$-\frac{1}{150}$	1	0	$\langle\frac{1}{3}\rangle$	0	0
0	0	1	0	0	0	1	1
$-\frac{7}{4}$	330	$\frac{1}{50}$	0	0	-2	0	0

对第五表再进行一次取主运算即得初始表. 这样经过六次迭代又回复到原来的状况,这就叫做循环. 注意到迭代前后的目标函数值都是零,且由(2.37)中给出的可行解对应的目标函数值是负的,这样下去永远得不到最优解.

如果一个基本可行解是退化的,那么有这种可能性:将 a_k 进基而使 a_{j_l} 离基时,有 $y_{l0}/y_{lk}=0$. 这意味着新的基本可行解也是退化的,并且新旧目标值相同. 于是理论上可能产生循环. 本节给出的摄动理论就是为了解决这个问题的,但其本身也有独立的意义.

考虑标准型线性规划

(LP)
$$\begin{aligned} &\min\ c^T x \\ &\text{s.t.}\ Ax=b \\ &\qquad x\geqslant 0 \end{aligned}$$

其中 $A=(a_1,a_2,\cdots,a_n)$. 对 $Ax=b$ 定义摄动方程组

$$Ax=b(\varepsilon),\ \varepsilon>0$$

其中

$$b(\varepsilon)=b+\varepsilon a_1+\varepsilon^2 a_2+\cdots+\varepsilon^n a_n \tag{2.38}$$

再考虑

$(\text{LP})_\varepsilon$
$$\begin{aligned} &\min\ c^T x \\ &\text{s.t.}\ Ax=b(\varepsilon) \\ &\qquad x\geqslant 0 \end{aligned}$$

定理2.10 设 $B=(a_1,a_2,\cdots,a_m)$为(LP)的可行基,则对于足够小的 $\varepsilon>0$,$(\text{LP})_\varepsilon$ 对应于基 B 的基本解是非退化的基本可行解.

证 $(\text{LP})_\varepsilon$ 关于基 B 的基本解为可行解,当且仅当

$$B^{-1}b(\varepsilon) \geqslant 0$$

利用单纯形表(2.19)中的记号以及(2.38)，$B^{-1}b(\varepsilon)$之分量为

$$y_{i0} + \varepsilon^i + \sum_{j=m+1}^{n} y_{ij}\varepsilon^j,\ i = 1,\cdots,m \tag{2.39}$$

由于 B 是(LP)的可行基，故

$$y_{i0} \geqslant 0,\ i = 1,\cdots,m \tag{2.40}$$

再由于(2.39)式第二项中 ε 的次数严格小于其第三项中任一个 ε 的次数，由此结合(2.40)知对充分小的 $\varepsilon>0$，(2.39)中的表达式严格为正，故结论为真.

定理2.11 存在 $\varepsilon_0>0$，使得对$\forall\ \varepsilon\in(0,\varepsilon_0)$，$(\mathrm{LP})_\varepsilon$ 是非退化的，即它的任一基本可行解都是非退化的.

证 任取 A 中的一个基，记对应的基变量的下标集为 $S=\{j_1,j_2,\cdots,j_m\}$(参见单纯形算法)，记非基变量的下标集为 T. 沿用对应于此基的单纯形表(2.25)中的记号，类似于(2.39)，现行的 $(\mathrm{LP})_\varepsilon$ 的单纯形表中 $b(\varepsilon)$列的内容为

$$b_i(\varepsilon) = y_{i0} + \varepsilon^{j_i} + \sum_{j\in T} y_{ij}\varepsilon^j,\ j_i \in S \tag{2.41}$$

上式中关于 ε 的非零多项式

$$y_{i0} + \varepsilon^{j_i} + \sum_{j\in T} y_{ij}\varepsilon^j,\ i = 1,\cdots,m \tag{2.42}$$

只有有限个零点. 故存在 $\varepsilon_0>0$使得(2.42)关于 ε 在$(0,\varepsilon_0)$中没有零点. 由于基的个数是有限的，故 $\varepsilon_0>0$可选择得使对所有的基变量的下标集 S，(2.41)中的 $b_i(\varepsilon)$在$(0,\varepsilon_0)$中没有零点. 这样，(2.41)表明对$\forall\ \varepsilon\in(0,\varepsilon_0)$，$(\mathrm{LP})_\varepsilon$ 的任一基本解或是不可行的，或是非退化的基本可行解. 证毕.

若(LP)有一个可行基 B 其由 A 的首 m 列组成，则由上两定理知对$\forall\ \varepsilon\in(0,\varepsilon_0)$，$(\mathrm{LP})_\varepsilon$ 是非退化的且至少有一个基本可行解. 从而可从基 B 出发对$(\mathrm{LP})_\varepsilon$ 关于某个 $\varepsilon\in(0,\varepsilon_0)$作单纯形算法. 若写出$(\mathrm{LP})_\varepsilon$ 的单纯形表，则除了 $b(\varepsilon)$列外，其余的列及判别数和(LP)关于同一个基的单纯形表是一样的. 在用单纯形法求解$(\mathrm{LP})_\varepsilon$ 时，若在某次迭代时有某个 $r_k<0$且 a_k 列的元皆非正，则由

定理2.6知$(LP)_\varepsilon$的目标值无下界，同理可知(LP)的目标值无下界(注意，此时有 $y_{i0}\geqslant 0, i=1,\cdots,m$)；余下的情况只能是得到了$(LP)_\varepsilon$的一个非退化的最优基本可行解且相应的判别数必全是非负的，对应地得到了(LP)的全是非负的判别数，由于上面的论述对$(0,\varepsilon_0)$中的 ε 的大小无关，以及此时在(2.41)中恒有 $b_i(\varepsilon)>0$，故必有 $y_{i0}\geqslant 0, i=1,\cdots,m$. 这样就得到了(LP)的一个最优基本可行解且相应的判别数皆非负. 从而，我们得到了

定理2.12 若线性规划(LP)有一个最优解，则这个规划必有一个最优基.

设(LP)的初始可行基由 A 的首 m 列组成. 对$(LP)_\varepsilon$的单纯形表作取主运算的过程进行简单的分析，可得对(LP)进行单纯形算法的一种选择主元的规则. 注意到对$(LP)_\varepsilon$的任一个现行的单纯形表，其 $b(\varepsilon)$ 列由(2.41)之右端给出，先利用表(2.25)将(2.41)改写成紧凑的形式：

$$b_i(\varepsilon) = y_{i0} + \sum_{j=1}^{n} y_{ij}\varepsilon^j,\ i = 1,\cdots,m \tag{2.43}$$

再设 a_k 进基，主元行 l 应取使下花括号内的表达式

$$\min_{i:y_{ik}>0}\left\{\frac{y_{i0}+\sum_{j=1}^{n} y_{ij}\varepsilon^j}{y_{ik}}\right\} \tag{2.44}$$

达到最小的 i(参见(2.8)式). 由于 $\varepsilon>0$ 可以足够小，故首选 l 为 I_0 中的元：

$$I_0=\{i: y_{i0}/y_{ik}=\min_{i:y_{ik}>0} y_{i0}/y_{ik}, i=1,\cdots,m \text{ 且 } y_{ik}>0\} \tag{2.45}$$

若 I_0 的元不唯一. 由(2.44)中花括号内 $i\in I_0$ 的元的大小依赖于 ε^j 的系数，故定义

$$I_j = \{i: y_{ij}/y_{ik} = \min_{i\in I_{j-1}} y_{ij}/y_{ik}, i\in I_{j-1}\}, j = 1,2,\cdots \tag{2.46}$$

直到某个 I_s 的元唯一. 设此元为 l，选 y_{lk} 为主元. 根据定理2.12以上的论述可知，这种选择主元的方法能避免循环(请读者自证 $s\leqslant m$). 称这种方法为“摄动法”或“字典顺序法”.

一个基本可行解若是非退化的，则其对应于唯一的基；若此解

为最优解，容易证明对应的基是最优基. 一个基本可行解若是退化的，则其可能对应于好几个基；若此解还是最优解，我们现在能够证明其对应的基中至少有一为最优基. 这样，我们得到定理2.12的一个加强，它可以看成定理2.3的逆. 我们把证明留给读者.

定理2.13 线性规划(LP)的任一个最优基本可行解都有一个最优基.

习 题

1. 假定对应于每一个非基变量 x_j 所对应的 j，都有判别数 $r_j>0$，试证相应的基本可行解是唯一的最优解.

2. 在单纯形法中，若不选择 r_j 为最负作为进基向量的标准，而选择

$$\max_i\{r_j\cdot y_{i0}/y_{ij}:y_{ij}>0\} \tag{2.47}$$

为最小作为进基向量的准则. 试证明这个准则能导致目标函数值的最大改进.

3. 用单纯形法求解

$$\begin{aligned}\max\ & 2x_1+4x_2+x_3+x_4\\ \text{s.t.}\ & x_1+3x_2+x_4\leqslant 4\\ & 2x_1+x_2\leqslant 3\\ & x_2+4x_3+x_4\leqslant 3\\ & x_i\geqslant 0,i=1,2,3,4\end{aligned}$$

4. 对上题的线性规划

(i) $b=(4,3,3)^T$ 中的元素能连续变动多少？使得原最优基不变？

(ii) $c=(2,4,1,1)^T$ 中的元素能连续变动多少，使得原最优基不变？

(iii) 对于 b 的很小的变化，最优值有什么变化？

(iv) 对于 c 的很小的变化，最优值有什么变化？

5. 用单纯形法求解

$$\begin{aligned}\max\ & 3x_1+x_2+3x_3\\ \text{s.t.}\ & 2x_1+x_2+x_3\leqslant 2\\ & x_1+2x_2+3x_3\leqslant 5\\ & 2x_1+x_2+x_3\leqslant 6\\ & x_1,\quad x_2,\quad x_3\geqslant 0\end{aligned}$$

6. 用两阶段单纯形法解

$$
\begin{aligned}
\min\ & -3x_1 + x_2 + 3x_3 - x_4 \\
\text{s.t.}\ & x_1 + 2x_2 - x_3 + x_4 = 0 \\
& 2x_1 - 2x_2 + 3x_3 + 3x_4 = 9 \\
& x_1 - x_2 + 2x_3 - x_4 = 6 \\
& x_i \geqslant 0,\ i = 1,\ 2,\ 3,\ 4
\end{aligned}
$$

7. 线性不等式组

$$Ax \geqslant b,\ x \geqslant 0,\ \text{其中}\ b \geqslant 0$$

可以化为标准型

$$Ax - Iy = b,\ x \geqslant 0,\ y \geqslant 0 \tag{2.48}$$

用下列方法先变换(2.48)：令 $b_k = \max_i\{b_i\}$，其中 b_i 是 b 的分量；将第 k 行分别减去其余各行作为新的对应的行．这样便得到新(2.48)．证明为获(2.48)的初始基本可行解，只要对新(2.48)附加单个的人工变量即可．并利用这一方法求下面不等式组在标准型下的一组基本可行解．

$$
\begin{aligned}
x_1 + 2x_2 + x_3 &\geqslant 4 \\
2x_1 + x_2 + x_3 &\geqslant 5 \\
2x_1 + 3x_2 + 2x_3 &\geqslant 6 \\
x_i \geqslant 0,\ i &= 1,2,3
\end{aligned}
$$

8. 举例说明，对应于一个退化的最优基本可行解的某个单纯形表，其判别数不全是非负的．

9. 设线性规划(LP)有一个最优基本可行解，它是退化的．设已有对应于此解的某个基的单纯形表(2.25)．记

$$I = \{i : y_{i0} = 0, i = 1, \cdots, m\} \tag{2.49}$$

证明：对现行表中的任一判别数 $r_k : r_k < 0$，集合

$$\{y_{ik} : y_{ik} > 0, i \in I\} \tag{2.50}$$

是非空的．由此再利用(2.45)和(2.46)及摄动理论证明定理2.13．

10. 解线性规划的大 M 法．考虑

$$
\begin{aligned}
\min\ & c^T x \\
\text{s.t.}\ & Ax = b \\
& x \geqslant 0
\end{aligned} \tag{2.51}
$$

设(2.51)有最优解．再考虑

$$
\begin{aligned}
\min\ & c^T x + Me^T y \\
\text{s.t.}\ & Ax + Iy = b \\
& x \geqslant 0, y \geqslant 0
\end{aligned} \tag{2.52}
$$

其中 $e=(1,1,\cdots,1)^T\in R^m$. 证明:存在 $m_0>0$,只要 $M\geqslant m_0$,就有

(i) 若 $\bar{x}$ 为(2.51)的最优基本可行解,则$(\bar{x},0)$为(2.52)的最优基本可行解;

(ii) 若$(\hat{x},\hat{y})$为(2.52)的最优基本可行解,则必有 $\hat{y}=0$,从而 $\hat{x}$ 为(2.51)的最优基本可行解.

11. 用摄动法求解例2.9.

12. 用单纯形理论结合摄动法证明第一章习题5的结果.

第三章 对偶理论

联系着线性规划的一个重要而有趣的问题是对偶线性规划. 对偶的概念首由 von Neumann 在1947年引入线性规划. 其后由 Gale, Kuhn, Tucker 三人表成精确的形式.

§3.1 对偶线性规划

我们先描述如下的对偶：

$$
\begin{array}{ll}
\text{原始问题} & \text{对偶问题} \\
\min\ c^T x & \max\ b^T \lambda \\
\text{s.t.}\ Ax \geqslant b & \text{s.t.}\ A^T \lambda \leqslant c \\
\quad x \geqslant 0 & \quad \lambda \geqslant 0
\end{array}
\tag{3.1}
$$

其中变量 λ 为 m 维列向量. 这里原始问题用不等式约束而不用等式约束，因为这样最具有对称性. (3.1)的两个问题也称为对称对偶. 给定了(3.1)的原问题，建立它的对偶的规则是这样的：

(i) 对偶问题用且仅用到原始问题的数据，即 A,b,c；

(ii) 一个问题的变量为 n 维，则另一个问题的变量为 m 维，它们的维数分别对应于矩阵 A 的列数和行数，且变量都要求非负；

(iii) 两个问题关于矩阵 A 的不等号是反向的，且对应于不等号"$\geqslant$"的规划其目标函数取极小，同时另一问题取极大.

线性规划的对偶具有对合性，即对偶问题的对偶是原问题. 请读者自行验证.

依照上述规则可以建立任何线性规划问题的对偶. 例如，考虑标准型线性规划

$$
\begin{aligned}
&\min\ c^T x \\
&\text{s.t.}\ Ax = b \\
&\qquad x \geqslant 0
\end{aligned}
$$

先把它化成等价的形式

$$\min\ c^T x$$
$$\text{s.t.}\ \begin{pmatrix} A \\ -A \end{pmatrix} x \geqslant \begin{pmatrix} b \\ -b \end{pmatrix}$$
$$x \geqslant 0$$

上面的规划形如(3.1)的原问题. 以 $u, v \in R^m$ 作为对偶变量，再依上述规则得对偶问题为

$$\max\ b^T u - b^T v$$
$$\text{s.t.}\ A^T u - A^T v \leqslant c$$
$$u \geqslant 0, v \geqslant 0$$

令 $\lambda = u - v$，则得到简化了的对偶如下：

原始问题	对偶问题	
$\min\ c^T x$	$\max\ b^T \lambda$	
$\text{s.t.}\ Ax = b$	$\text{s.t.}\ A^T \lambda \leqslant c$	(3.2)
$x \geqslant 0$		

这是一种非对称形式的对偶. 注意，此时的对偶变量 λ 是自由变量.

例3.1 考虑例1.1中的线性规划，它是(3.1)中的对偶问题的形式. 根据对偶之对合性，它的对偶是(3.1)中原问题的形式，因此为

$$\begin{aligned} \min\ & 3500y_1 + 1500y_2 + 10000y_3 \\ \text{s.t.}\ & y_1 + y_2 + 5y_3 \geqslant 5 \\ & y_1 \qquad + 2y_3 \geqslant 3 \\ & y_1, y_2, y_3 \geqslant 0 \end{aligned} \tag{3.3}$$

对照例1.1中的表，对 $i=1,2,3$，令

y_i：使用第 i 种原料一个单位后所得之利润.

则(3.3)的目标函数表所有原料用完后所得之总利润；而前两个不等式则表示了由例1.1进行的生产能给予的有关利润的基本保证. 这个问题对原料商确定原料的销售价也很有参考价值(此时，y_i 应定义为第 i 种原料的单位销售差价).

例3.2（运输问题的对偶） 第一章中所描述的运输问题形如(3.2)的原问题，因此它的对偶也就形如(3.2)的对偶问题. 由于运输问题的有关矩阵很特殊，我们可以将其对偶问题表达得具体些. 写出运输问题的矩阵为（其中未列出的元皆为零）

$$
\begin{array}{c}
\\
m\text{ 行}\left\{\begin{array}{c} \\ \\ \\ \\ \end{array}\right. \\
n\text{ 行}\left\{\begin{array}{c} \\ \\ \\ \\ \end{array}\right.
\end{array}
\begin{array}{c}
\begin{array}{ccccccccccccc}
x_{11} & x_{12} & \cdots & x_{1n} & x_{21} & x_{22} & \cdots & x_{2n} & \cdots & x_{m1} & x_{m2} & \cdots & x_{mn}
\end{array}\\
\begin{pmatrix}
1 & 1 & \cdots & 1 & & & & & & & & & \\
 & & & & 1 & 1 & \cdots & 1 & & & & & \\
 & & & & & & & & \ddots & & & & \\
 & & & & & & & & & 1 & 1 & \cdots & 1 \\
1 & & & & 1 & & & & \cdots & 1 & & & \\
 & 1 & & & & 1 & & & & & 1 & & \\
 & & \ddots & & & & \ddots & & & & & \ddots & \\
 & & & 1 & & & & 1 & \cdots & & & & 1
\end{pmatrix}
\end{array}
$$

以及

$$\bar{b} = (a_1, a_2, \cdots, a_m, b_1, b_2, \cdots, b_n)^T$$

$$c = (c_{11}, c_{12}, \cdots, c_{1n}, c_{21}, c_{22}, \cdots, c_{2n}, \cdots, c_{m1}, c_{m2}, \cdots, c_{mn})^T$$

再记对偶变量

$$\lambda = (u_1, u_2, \cdots, u_m, v_1, v_2, \cdots, v_n)^T$$

代入(3.2)的对偶问题，我们得到运输问题的对偶的具体表达式

$$
\begin{aligned}
&\max \sum_{i=1}^{m} a_i u_i + \sum_{j=1}^{n} b_j v_j \\
&\text{s.t.}\ \ u_i + v_j \leqslant c_{ij} \\
&\qquad i = 1, 2, \cdots, m;\ j = 1, 2, \cdots, n
\end{aligned}
\tag{3.4}
$$

注意，运输问题的对偶的形式特别简洁.

§3.2 对偶定理

应该指出，以上的对偶问题的建立并不是随意的. 本节给出以下一些深刻的结果. 这些结果可以看成是对定义的一种启发. 同时更重要的是导致新的算法.

以下，我们只研究标准型线性规划

$$\begin{aligned} &\min\ c^T x \\ &\text{s.t.}\ Ax = b \\ &\qquad x \geqslant 0 \end{aligned} \tag{3.5}$$

及其相应的对偶问题

$$\begin{aligned} &\max\ b^T\lambda \\ &\text{s.t.}\ A^T\lambda \leqslant c \end{aligned} \tag{3.6}$$

我们立即有下面的重要关系.

引理3.3 若 x 和 λ 分别为(3.5)和(3.6)的可行解，则有

$$b^T\lambda \leqslant c^T x \tag{3.7}$$

证 利用关系式 $Ax=b, A^T\lambda\leqslant c, x\geqslant 0$立得

$$\lambda^T b = \lambda^T(Ax) = (A^T\lambda)^T x \leqslant c^T x$$ 证毕.

上面的引理一般称为弱对偶.

引理3.4 若 x^0和 λ^0分别为(3.5)和(3.6)的可行解，且满足

$$c^T x^0 = b^T\lambda^0 \tag{3.8}$$

则 x^0和 λ^0分别为各自对应的问题的最优解.

证 对(3.5)的任一可行解 x，由(3.7)得

$$c^T x \geqslant b^T\lambda^0 = c^T x^0$$

这表明 x^0是(3.5)的最优解. 同理可得另一结论.

引理3.5 若 B 是(3.5)的最优基，则

$$\lambda^0 = (c_B^T B^{-1})^T \tag{3.9}$$

为(3.6)的最优解，且(3.5)和(3.6)有相同的最优值.

证 由 B 为(3.5)的最优基，按定义有

$$c^T - c_B^T B^{-1} A \geqslant 0$$

因此，$\lambda^0=(c_B^T B^{-1})^T$ 对(3.6)是可行的. 设 x^0是(3.5)的关于 B 的基本可行解，则由单纯形表(2.35)即知

$$c^T x^0 = c_B^T B^{-1} b = (\lambda^0)^T b$$

于是，结论由引理3.4立得. 证毕.

在对(3.5)施行单纯形算法的任一过程中，我们都可以形成向量$(c_B^TB^{-1})^T$(参见修正单纯形算法中的说明(ii))，这个向量称为单纯形乘子.若这个向量又是对偶可行的(即满足(3.6)中的不等式)，注意到单纯形法中总有$B^{-1}b\geqslant 0$，结合定理2.3和引理3.5可知：$B^{-1}b$对应着(3.5)的一个最优基本可行解，而$(c_B^TB^{-1})^T$则是(3.5)的对偶问题(3.6)的一个最优解.这表明在求得一个问题的最优解的同时也获得了其对偶问题的最优解.

引理3.6 对于线性规划(3.5)和(3.6)，若两规划均有可行解，则它们都有最优解，而且它们的最优值相同.

证 若两规划均有可行解，则由引理3.3知(3.5)的目标函数在可行集上有下界，从而它有一个最优基(参见第一章习题5和定理2.12).再利用引理3.5知结论为真.

将以上各引理归纳起来得如下的几个重要结果.

定理3.7(对偶定理) 对于线性规划(3.5)和(3.6)，有

(i) 若其中之一有最优解，则另一规划也有最优解，且两规划之最优值相同；

(ii) 若其中之一目标函数在可行集上无界，则另一规划无可行解；

(iii) 若其中之一无可行解，则另一规划无可行解或目标函数在可行集上无界.

注 这里的"目标函数在可行集上无界"指的是：对求极小的问题言"在可行集上无下界"，对求极大的问题言"在可行集上无上界".

证 依据对偶之对合性，只须证若条件对(3.5)成立则结论对(3.6)成立.

(i) 即引理3.5和定理2.12的结合；

(ii) 由引理3.3立得；

(iii) 此款的逆否命题为：另一规划有可行解且目标函数在可行集上有界，则原规划有可行解；而条件意味着所论规划有最优解(从而它也是可行解)，故此逆否命题为(i)之推论.证毕.

定理3.8(松紧定理)　若 x^0 和 λ^0 分别是(3.5)和(3.6)的可行解,则它们分别为对应问题的最优解的充要条件为

$$(c - A^T\lambda^0)^T x^0 = 0 \tag{3.10}$$

证　按引理3.4和引理3.6可知,x^0 和 λ^0 分别为对应问题的最优解的充要条件为

$$c^T x^0 = b^T\lambda^0 = (Ax^0)^T\lambda^0$$

由此即推得(3.10).证毕.

仔细分析(3.10),因有 $c-A^T\lambda^0 \geqslant 0$ 及 $x^0 \geqslant 0$,(3.10)等价于

$$(c_j - (\lambda^0)^T a_j)x_j^0 = 0,\ j = 1,\cdots,n$$

或等价于

$$\begin{cases} x_j^0 > 0 \Rightarrow (\lambda^0)^T a_j = c_j \\ (\lambda^0)^T a_j < c_j \Rightarrow x_j^0 = 0,\ j = 1,\cdots,n \end{cases} \tag{3.11}$$

上式称为松紧关系式.因为有

$$(\lambda^0)^T a_j \leqslant c_j,\ j = 1,\cdots,n$$

及

$$x_j^0 \geqslant 0,\ j = 1,\cdots,n$$

(3.11)表明:若 j 对一个问题是松的(在≤或≥的关系式中取严格不等号),则此 j 对对偶问题是紧的(在≤或≥的关系式中取等号).有人称关系式(3.11)为互补松弛.

将定理3.8换一个说法,便得

定理3.9　$x \in R^n$ 为(3.5)的最优解的充要条件是:存在 $\lambda \in R^m$,使得

$$\begin{cases} A^T\lambda - c \leqslant 0 \\ (A^T\lambda - c)^T x = 0 \\ Ax = b \\ x \geqslant 0 \end{cases}$$

这个定理表明线性规划问题的求解可以转化为对不等式组的求解,它对于某些非线性问题的转化也能扮演重要角色.

关于对称对偶(3.1)式,其对偶定理有与定理3.7完全相同的

叙述;其松紧定理及应用则放在习题中讨论.

§3.3 对偶单纯形法

前面我们已经说了,给定一个线性规划,它的对偶问题是可以由原始规划的数据表示出来的;而对偶定理告诉我们,知道了原始问题或对偶问题的最优解,就能得到另一个问题的最优解而无须对它另外求解.所以,如果对偶问题在计算上更吸引人的话,就可以通过求解对偶问题来求解原始问题.

不是去构造对偶问题的单纯形表,而是在原始问题的表上进行对偶处理.依据这个思想的完整的技巧便是对偶单纯形法.

仍考虑标准型线性规划

$$\text{(LP)}\qquad \begin{aligned} \min\ & c^T x \\ \text{s.t.}\ & Ax=b \\ & x\geqslant 0 \end{aligned}$$

设已给的基为 $B=(a_{j_1},a_{j_2},\cdots,a_{j_m})$,其中 $a_{j_i},i=1,2,\cdots,m$ 是 A 的 m 个线性无关列.对应地记 $c_B=(c_{j_1},c_{j_2},\cdots,c_{j_m})^T$,$x_B=(x_{j_1},x_{j_2},\cdots,x_{j_m})^T$.对应的单纯形表为

$$\begin{bmatrix} B^{-1}A & B^{-1}b \\ c^T-c_B^TB^{-1}A & -c_B^TB^{-1}b \end{bmatrix}$$

单纯形算法是:给定一个可行基 B,即满足 $B^{-1}b\geqslant 0$.通过取主运算,使得在保持可行性的条件下(即表中除去右下角元的最后列恒非负)最后的判别数皆非负.现在提出一个相反的求解方法:如果给定的基 B 是使得判别数非负,即满足 $c^T-c_B^TB^{-1}A\geqslant 0$,但未必有 $B^{-1}b\geqslant 0$.能否通过取主运算,使得在判别数恒为非负的条件下,最后的表中除右下角元外的最后列皆非负?回答是肯定的.这样的方法的实现便是对偶单纯形法.为了看得更清楚些,我们给出

定义3.10 对于给定的基 B,若

$$\lambda=(c_B^TB^{-1})^T \tag{3.12}$$

对于(LP)的对偶问题是可行的,即满足 $c^T-\lambda^TA\geqslant 0$,则称 B 为对

偶可行基.

这样，对于对偶单纯形法，从原始问题的角度来看，它的做法是保持最末行的最优性条件向实现可行性作运算. 然而从对偶的角度来看，它是在向最优性进行的同时保持可行性. 使用的手段还是取主运算.

注意，对应于任一现行的对偶可行基，对偶问题的目标函数值是 $\lambda^T b = c_B^T B^{-1} b$. 它仍对应于表的右下角元.

现在我们来推导方法. 首先，对于给定的对偶可行基 B，据定义有 $c^T - \lambda^T A \geqslant 0$，其中 λ 由(3.12)给出，于是有

$$\begin{cases} c_{j_i} - \lambda^T a_{j_i} = 0,\ i = 1,\cdots,m \\ c_j - \lambda^T a_j \geqslant 0,\ j = 1,\cdots,n \end{cases} \tag{3.13}$$

如果 $x_B = B^{-1} b \geqslant 0$，则已得原始问题的最优解.

其次，设 x_B 有某个分量为负，记

$$x_{B_l} = \min_i \{x_{j_i}\} < 0 \tag{3.14}$$

选择 l 作为主元行. 下面考虑如何选择主元列 k，使得取主运算能够保证

(i) 对偶目标函数值有所上升，即 $\lambda^T b \leqslant \bar{\lambda}^T b$，$\bar{\lambda}$ 为迭代后的对偶可行解；

(ii) 新基 $\bar{B}$ 仍然是对偶可行基，即对于 $\bar{\lambda}^T = c_{\bar{B}}^T \bar{B}^{-1}$，有

$$c^T - \bar{\lambda}^T A \geqslant 0.$$

用 u^l 表示 B^{-1} 的第 l 行，设想

$$\bar{\lambda}^T = \lambda^T - \varepsilon u^l \tag{3.15}$$

其中 ε 为一待定的数. 于是

$$\bar{\lambda}^T b = \lambda^T b - \varepsilon u^l b = \lambda^T b - \varepsilon x_{B_l} \tag{3.16}$$

为了满足条件(i)，必须且仅需 $\varepsilon \geqslant 0$. 特别若 $\varepsilon > 0$，则对偶目标函数值严格增加.

为了保持对偶可行性，由(ii)必须有

$$c^T - \bar{\lambda}^T A \geqslant 0 \tag{3.17}$$

将(3.15)代入(3.17)整理得

$$\varepsilon u^l A \geqslant \lambda^T A - c^T$$

按分量写出来就是

$$\varepsilon u^l a_j \geqslant \lambda^T a_j - c_j,\ j = 1,\cdots,n \tag{3.18}$$

注意到 $u^l a_j = y_{lj}$，$\lambda^T a_j = z_j$，(3.18)即为

$$\varepsilon y_{lj} \geqslant z_j - c_j,\ j = 1,\cdots,n \tag{3.19}$$

其中 y_{lj}以及 $c_j - z_j(\geqslant 0)$，$j=1,\cdots,n$ 就是现行单纯形表中的有关数据.

若 $y_{lj}\geqslant 0$，$j=1,\cdots,n$，则由(3.19)对任意的 $\varepsilon>0$，都能保证 $\bar{\lambda}$ 的对偶可行性. 此时由(3.16)可知相应的对偶目标函数值趋于正无穷. 由对偶定理，原始问题无可行解.

若 y_{lj}中有为负者. 若要保持对偶可行性，只要取

$$\varepsilon_0 = \frac{z_k - c_k}{y_{lk}} = \min_{y_{lj}<0}\left\{\frac{z_j - c_j}{y_{lj}}\right\} \tag{3.20}$$

(3.14)和(3.20)便是定主元行 l 和主元列 k 的准则.

最后要验证的是对单纯形表进行取主运算后所得到的 $\bar{\lambda} = (\bar{c}_B^T \bar{B}^{-1})^T$就是由(3.15)和(3.20)所定义的 $\bar{\lambda}$，即证

$$\bar{c}_B^T \bar{B}^{-1} = \lambda^T - \varepsilon_0 u^l = c_B^T B^{-1} + \frac{c_k - z_k}{y_{lk}} u^l \tag{3.21}$$

这里

$$c_B = (c_{j_1}, c_{j_2}, \cdots, c_{j_{l-1}}, c_{j_l}, c_{j_{l+1}}, \cdots, c_{j_m})^T$$

$$\bar{c}_B = (c_{j_1}, c_{j_2}, \cdots, c_{j_{l-1}}, c_k, c_{j_{l+1}}, \cdots, c_{j_m})^T$$

$$\bar{B}^{-1} = E_{lk} B^{-1}$$

其中 E_{lk}见(2.36)，它是取主运算的矩阵. 于是

$$\bar{c}_B^T \bar{B}^{-1} = \bar{c}_B^T E_{lk} B^{-1}$$

$$= \left(c_{j_1}, c_{j_2}, \cdots, c_{j_{l-1}}, \frac{c_k - \sum\limits_{i=1, i\neq l}^{m} y_{ik} c_{j_i}}{y_{lk}}, c_{j_{l+1}}, \cdots, c_{j_m} \right) B^{-1}$$

$$= c_B^T B^{-1} + \frac{c_k - \sum\limits_{i=1}^{m} y_{ik} c_{j_i}}{y_{lk}} u^l$$

$$=c_B^T B^{-1}+\frac{c_k-z_k}{y_{lk}}u^l$$

因为迭代中只是进行了一个基本变量与另一个非基变量的交换，所以新的判别数行仍有 m 个为零的元. 且根据以上取主规则，其它的 $n-m$ 个判别数非负.

我们将上面的推导总结成如下的

对偶单纯形算法 对于(LP)的一个对偶可行基

$$B=(a_{j_1},a_{j_2},\cdots,a_{j_m})$$

第一步 定主元行 l

$$x_B=B^{-1}b=(x_{j_1},x_{j_2},\cdots,x_{j_m})^T$$

$$x_{j_l}=\min_i\{x_{j_i}\}$$

若 $x_{j_l}\geqslant 0$，则算法结束，对应于 x_B 的解是原问题的最优解. 否则

第二步 定主元列 k

$$\lambda=(c_B^T B^{-1})^T$$

$$\frac{\lambda^T a_k-c_k}{y_{lk}}=\min_{y_{lj}<0}\left\{\frac{\lambda^T a_j-c_j}{y_{lj}}\right\}$$

若无 k，即 $y_{lj}\geqslant 0,j=1,\cdots,n$，则算法结束，原问题无可行解. 否则

第三步 用 a_k 代替 a_{j_l} 形成新基 B，变换标号

$$B^{-1}\Leftarrow E_{lk}B^{-1}$$

$$c_{j_l}\Leftarrow c_k$$

$$j_l\Leftarrow k$$

返回第一步.

关于算法的收敛性证明可参见§2.2.

例3.11 考虑问题

$$\begin{aligned}
&\min\ 3x_1+4x_2+5x_3\\
&\text{s.t.}\ \ x_1+2x_2+3x_3\geqslant 5\\
&\qquad 2x_1+2x_2+x_3\geqslant 6\\
&\qquad x_i\geqslant 0,\ i=1,2,3
\end{aligned}$$

通过引进剩余变量及改变不等号的方向得

初 始 表

−1	−2	−3	1	0	−5
(−2)	−2	−1	0	1	−6
3	4	5	0	0	0

现有一个现成的基，它由剩余变量组成，而且是对偶可行的. 于是即可用对偶单纯形法. 取 x_B 的最负分量作为主元行，即上表对应于−6的第二行，计算比

$$3/(-2)=-3/2,\quad 4/(-2)=-2,\quad 5/(-1)=-5$$

取最大的负数作为主元列，即上表之第一列. 主元已在上表括出. 作取主运算，包括表上最末行相应的变换，它是使对应于新基的判别数为零. 这样得

第 一 表

0	(−1)	−5/2	1	−1/2	−2
1	1	1/2	0	−1/2	3
0	1	7/2	0	3/2	−9

主元行为第一行，计算比

$$1/(-1)=-1,\quad (7/2)/(-5/2)=-7/5,$$
$$(+3/2)/(-1/2)=-3.$$

取最大的负数，得主元列为第二列. 这样又得

第 二 表

0	1	5/2	−1	1/2	2
1	0	−2	1	−1	1
0	0	1	1	1	−11

第二表产生原始问题的一个最优解(1,2,0).

在对偶可行基容易得到的情况下，可选用对偶单纯形法. 以下列出一些常见的情况：

(i) 求解

$$\begin{aligned} &\min\ c^T x \\ &\text{s.t.}\ Ax \geqslant b \\ &\quad\quad x \geqslant 0 \end{aligned}$$

其中 $c \geqslant 0$. 这时，我们有单纯形表中现成的对偶可行基 I：

$$\begin{pmatrix} -A & I & -b \\ c^T & 0 & 0 \end{pmatrix}$$

可直接进行对偶单纯形计算. 上例即如此.

(ii) 对于求解如下的一组规划问题

$$(\text{LP}^t) \quad \begin{aligned} &\min\ c^T x \\ &\text{s.t.}\ Ax = b^{(t)} \\ &\quad\quad x \geqslant 0 \end{aligned}$$

$t=1,2,\cdots,k$. 这时可先用单纯形法求得(LP^1)的一个最优基 B，从而 B 就是(LP^t)的对偶可行基. 于是可用对偶单纯形法求解(LP^t)，$t=2,3,\cdots,k$.

(iii) 设我们已求得(LP)的一个最优基 B. 然后发现需要增添某个约束条件 $\alpha^T x \leqslant d$，即要求解

$$\begin{aligned} &\min\ c^T x \\ &\text{s.t.}\ Ax = b \\ &\quad\quad \alpha^T x + x_{n+1} = d \\ &\quad\quad x \geqslant 0,\ x_{n+1} \geqslant 0 \end{aligned}$$

这时我们若取新问题的基为

$$\bar{B} = \begin{bmatrix} B & 0 \\ \alpha_B^T & 1 \end{bmatrix} \tag{3.22}$$

其中 α_B 为 α 中对应于 B 的列的各元组成，1是 x_{n+1} 的系数，则 $\bar{B}$ 就是新问题的对偶可行基. 读者可自行验证之.

§3.4 参数线性规划

对于标准型线性规划

(LP)
$$\begin{aligned} \min\ & c^T x \\ \text{s.t.}\ & Ax=b \\ & x\geqslant 0 \end{aligned}$$

我们知道，其中的 c,A,b 都是常量. 若考虑到它们有时会变动，则得到了如下的线性规划：

(LP_θ)
$$\begin{aligned} \min\ & c(\theta)^T x \\ \text{s.t.}\ & A(\theta)x=b(\theta) \\ & x\geqslant 0 \end{aligned}$$

其中 $c(\theta),A(\theta),b(\theta)$ 是 θ 的函数. 对应于任一个 θ 得一个线性规划 (LP_θ)，称 θ 为参数，称 (LP_θ) 为参数线性规划. 为简单起见，设 $\theta\in R$. 现在的问题是：对怎样的 θ，(LP_θ) 有最优解 $x(\theta)$？θ 的变化对最优解及最优值的影响如何？在最优解 $x(\theta)$ 存在的时候还有一个怎样求解的问题.

对参数规划的研究来源于实际问题的需要. 比如说，在运输问题中单位货物的运价不是一成不变的（对应于 c 的变化）；物资的收发量也可能会变动（对应于 b 的变化），等等. 因此，对参数规划的研究有其实用意义. 其理论研究上的意义也日显重要.

对一般参数线性规划的研究比较困难；本节只考虑较为简单的情况.

1. 目标函数含参数的线性规划

考虑问题

(LPC_θ)
$$\begin{aligned} \min\ & c(\theta)^T x \\ \text{s.t.}\ & Ax=b \\ & x\geqslant 0 \end{aligned}$$

其中 $c(\theta)=(c_1(\theta),c_2(\theta),\cdots,c_n(\theta))^T$.

定义3.12 设对某个 j 有

$$\alpha c_j(\theta_1)+(1-\alpha)c_j(\theta_2)\leqslant c_j(\alpha\theta_1+(1-\alpha)\theta_2)$$
$$\forall\ \alpha\in[0,1],\forall\ \theta_1,\theta_2\in R \qquad (3.23)$$

则称 $c_j(\theta)$ 为 R 上的凹函数；若(3.23)对 $j=1,\cdots,n$ 都成立，则称

$c(\theta)$为 R 上的凹函数.

关于凹函数的一些性质可参见第七章.

引理3.13 设 $c(\theta)$是 R 上的凹函数. 若(LPC_θ)对 θ_1和 θ_2都有最优解，$\theta_1 \leqslant \theta_2$，则($\mathrm{LPC}_\theta$)对 $\theta \in [\theta_1, \theta_2]$都有最优解.

证 $\hat{\theta} \in [\theta_1, \theta_2]$当且仅当存在 $\alpha \in [0,1]$，使得

$$\hat{\theta} = \alpha\theta_1 + (1-\alpha)\theta_2 \tag{3.24}$$

设 x^s 为(LPC_{θ_s})的最优解，$s=1,2$，易知

$$\alpha x^1 + (1-\alpha)x^2$$

是($\mathrm{LPC}_{\hat{\theta}}$)的可行解. 今写出($\mathrm{LPC}_\theta$)的对偶问题：

(LDC_θ) $\quad \max\ b^T\lambda$

$$\text{s.t.}\ A^T\lambda \leqslant c(\theta)$$

因为(LPC_{θ_s})有最优解，由定理3.7知(LDC_{θ_s})亦有最优解，$s=1,2$. 设它们的最优解分别为 λ^1和 λ^2，于是就有

$$A^T\lambda^1 \leqslant c(\theta_1) \tag{3.25}$$

$$A^T\lambda^2 \leqslant c(\theta_2) \tag{3.26}$$

分别以 α 和$1-\alpha$ 乘以(3.25)和(3.26)，然后将所得之结果相加，再利用(3.23)，得

$$A^T(\alpha\lambda^1 + (1-\alpha)\lambda^2) \leqslant \alpha c(\theta_1) + (1-\alpha)c(\theta_2)$$
$$\leqslant c(\alpha\theta_1 + (1-\alpha)\theta_2)$$

由(3.24)知，上式表明 $\alpha\lambda^1+(1-\alpha)\lambda^2$是($\mathrm{LDC}_{\hat{\theta}}$)的可行解.

由此，($\mathrm{LPC}_{\hat{\theta}}$)和($\mathrm{LDC}_{\hat{\theta}}$)都有可行解，利用引理3.6知($\mathrm{LPC}_{\hat{\theta}}$)有最优解. 证毕.

上面的论证法是巧妙地运用对偶理论的范例.

引理3.14 设 $c(\theta)$是 R 上的连续函数，设有数列$\{\theta_s\}_{s=1}^{\infty}$满足 $\theta_s \to \hat{\theta}$ 且 $\hat{\theta}$ 有限. 若(LPC_{θ_s})有最优解，$s=1,2,\cdots$，则($\mathrm{LPC}_{\hat{\theta}}$)亦有最优解.

证 因(LPC_{θ_s})有最优解，由定理2.13知其有最优基，任取定一个最优基 B_s，$s=1,2,\cdots$；又因为(LPC_{θ_s})的基的个数与 s 无关且有限，故$\{B_s\}_{s=1}^{\infty}$中必有某个 B_{s_0}重复无限次，记为 B. 不失一般性，

与 B 对应的 $\{s\}$ 的子列仍记 $\{s\}$. 于是，$(\mathrm{LPC}_{\theta_s})$ 的关于最优基 B 的判别数应满足

$$c(\theta_s)^T - c_B(\theta_s)^T B^{-1} A \geqslant 0 \tag{3.27}$$

利用 $c(\theta)$ 的连续性由上式即得 $c(\hat{\theta})^T - c_B(\hat{\theta})^T B^{-1} A \geqslant 0$. 这说明 B 是 $(\mathrm{LPC}_{\hat{\theta}})$ 的最优基. 证毕.

由于 R 上的凹函数是局部 Lipschitz 连续的(参见第七章定理 7.5)，结合上两引理可得

定理3.15 对参数规划 (LPC_θ)，设 $c(\theta)$ 是 R 上的凹函数，设 (LPC_θ) 对某个 θ_0 有最优解，则存在某个包含 θ_0 的闭区间 $[\underline{\theta},\overline{\theta}]$，满足

(i) (LPC_θ) 对 $\theta\in[\underline{\theta},\overline{\theta}]$ 有最优解，对 $\theta<\underline{\theta}$ 及 $\theta>\overline{\theta}$ 无最优解；

(ii) $v(\theta)$ 是 $[\underline{\theta},\overline{\theta}]$ 上的局部 Lipschitz 连续函数，其中 $v(\theta)$ 为 (LPC_θ) 的最优值.

证 (i)考虑集合

$$G_1 = \{\theta:\theta \geqslant \theta_0, (\mathrm{LPC}_\theta) \text{ 有最优解}\}$$

因为 $(\mathrm{LPC}_{\theta_0})$ 有最优解，故 G_1 非空. 记 G_1 的上确界为 $\overline{\theta}$，则存在 $\{\theta_s\}\subseteq G_1$，$\theta_s\to\overline{\theta}$. 由 $\overline{\theta}$ 之定义知，(LPC_θ) 对 $\theta>\overline{\theta}$ 无最优解的结论平凡成立. 若 $\overline{\theta}$ 有限，由引理3.14知 $(\mathrm{LPC}_{\overline{\theta}})$ 有最优解，再由引理3.13知 (LPC_θ) 对 $\theta\in[\theta_0,\overline{\theta}]$ 都有最优解；若 $\overline{\theta}=+\infty$，类似地得 (LPC_θ) 对 $\theta\in[\theta_0,+\infty)$ 都有最优解. 再记

$$G_2 = \{\theta:\theta \leqslant \theta_0, (\mathrm{LPC}_\theta) \text{ 有最优解}\}$$

与上类似的论证便完成了(i)的证明.

(ii) 记 (LPC_θ) 的基本可行解集合为 $\{x^1,x^2,\cdots,x^q\}$. 由定理1.10之(ii)立得

$$v(\theta) = \min_{1\leqslant i\leqslant q}\{c(\theta)^T x^i\},\ \theta\in[\underline{\theta},\overline{\theta}] \tag{3.28}$$

由于每个 $c(\theta)^T x^i$ 都是 R 上的局部 Lipschitz 连续函数，故上式表明 $v(\theta)$ 是 $[\underline{\theta},\overline{\theta}]$ 上的局部 Lipschitz 连续函数. 证毕.

下面讨论定理中的 $\underline{\theta}$ 和 $\overline{\theta}$ 的求法，仍设 $c(\theta)$ 为 R 上的凹函数. 我们只给出 $\overline{\theta}$ 的求法，对 $\underline{\theta}$ 的求解可如法炮制. 设对 $(\mathrm{LPC}_{\theta_0})$ 已有了一

个最优基 B_0，写出(LPC_θ)关基 B_0的单纯形表

$$\begin{bmatrix} B_0^{-1}A & B_0^{-1}b \\ c(\theta)^T - c_{B_0}(\theta)^T B_0^{-1}A & -c_{B_0}(\theta)^T B_0^{-1}b \end{bmatrix} \tag{3.29}$$

表中已有 $B_0^{-1}b \geqslant 0$，再考虑关于 θ 的不等式组

$$c(\theta)^T - c_{B_0}(\theta)^T B_0^{-1}A \geqslant 0 \tag{3.30}$$

注意，(3.30)对 $\theta=\theta_0$是成立的；若某个 $\hat{\theta}$ 使(3.30)成立，则 B_0也是($LPC_{\hat{\theta}}$)的最优基. 设使(3.30)成立的最大的 θ 为$\bar{\theta}_0$，分若干情况：

(1)若$\bar{\theta}_0$不存在，即对任意 $M>\theta_0$总有 $\theta\geqslant M$ 使(3.30)成立，则$\bar{\theta}=+\infty$；

(2)若$\bar{\theta}_0$存在，即(3.30)对 $\theta=\bar{\theta}_0$成立，且存在(3.30)左端向量的第 k 个分量，记为 $r_k(\theta)$，其满足

$$r_k(\bar{\theta}_0)=0;\ r_k(\theta)<0,\ \theta>\bar{\theta}_0 \text{ 且 } \theta-\bar{\theta} \text{ 充分小} \tag{3.31}$$

注意，(3.30)中有 m 个不等式对任意 θ 恒作为等式成立，它们是关于基 B_0的基变量所对应的式子. 从而由(3.31)之第二式知对应的 a_k 是关于 B_0的非基向量；而 $r_k(\theta)$则是对应的判别数.

选 a_k 作为进基向量，分两种情况讨论.

(a) 在(3.29)中 a_k 列的元 y_{ik}，$i=1,\cdots,m$ 皆非正. 注意到诸 y_{ik}皆与 θ 无关及(3.31)之第二式，利用定理2.6知(LPC_θ)对充分接近$\bar{\theta}_0$的 $\theta>\bar{\theta}_0$无最优解；从而由引理3.13即推得(LPC_θ)对任意 $\theta>\bar{\theta}_0$亦无最优解. 因此$\bar{\theta}=\bar{\theta}_0$；否则，

(b) 取 $l: y_{l0}/y_{lk}=\min\limits_{y_{ik}>0}\{y_{i0}/y_{ik}\}$(若 l 的选取不唯一则使用字典顺序法)，取 y_{lk}为主元对(3.29)作取主运算得新基 B_1.

注意，(3.31)的第一式 $r_k(\bar{\theta}_0)=0$表明：在 $\theta=\bar{\theta}_0$时，表(3.29)的最末行在取主运算后未有变动；由于(3.30)对于 $\theta=\bar{\theta}_0$成立，故 B_0和 B_1都是($LPC_{\bar{\theta}_0}$)的最优基.

视 B_1为 B_0重复以上过程，有限次重复后即找到$\bar{\theta}$.

在

$$c(\theta)=c+\theta\bar{c}$$

时，其中 c 和 $\hat{c}$ 为给定的 n 维列向量，(LPC_θ)成为

$$(\mathrm{LP}\hat{\mathrm{C}}_\theta) \qquad \begin{aligned} &\min\ (c+\theta\hat{c})^T x \\ &\text{s.t.}\ \ Ax=b \\ &\qquad\ x\geqslant 0 \end{aligned}$$

注意，向量 $c+\theta\hat{c}$ 的任一分量自然满足(3.23)，故定理3.15对$(\mathrm{LP}\hat{\mathrm{C}}_\theta)$亦成立. 此时，(3.28)的花括号中有

$$(c+\theta\hat{c})^T x^i = (\hat{c}^T x^i)\theta + (c^T x^i) \triangleq l_i(\theta) \tag{3.32}$$

它是 θ 的线性函数. 仍记$(\mathrm{LP}\hat{\mathrm{C}}_\theta)$的最优值为 $v(\theta)$. 利用(3.28)我们能够给出在 $q=4$时的 $v(\theta)$的图像，其中粗线所示表 $v(\theta)$. 结合(3.32)可从图3.1中看出：在$[\underline{\theta},\theta_1]$中基本可行解 x^1是最优解，在$[\theta_1,\theta_2]$中基本可行解 x^2是最优解，在$[\theta_2,\overline{\theta}]$中基本可行解 x^3是最优解.

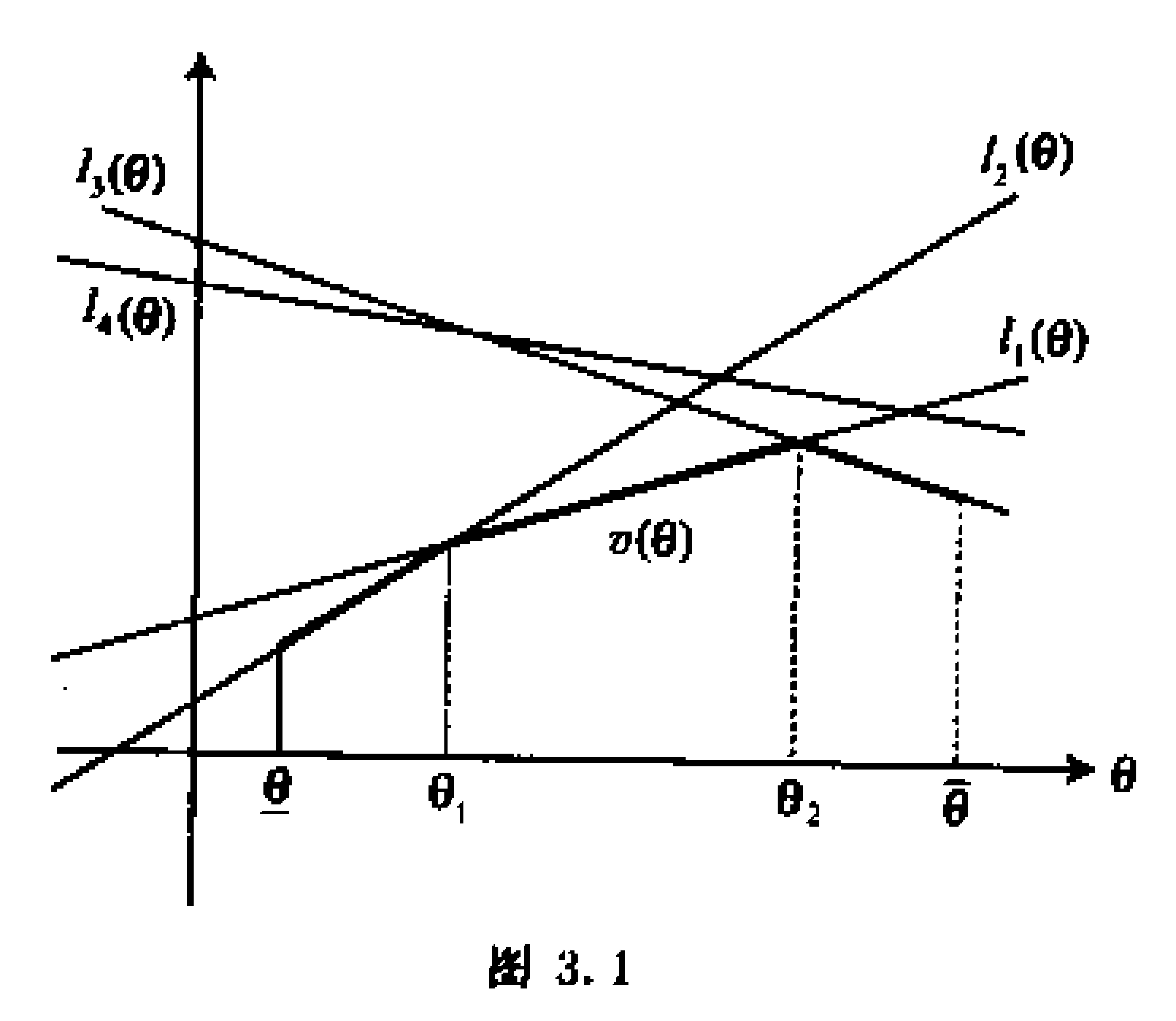


图 3.1

定理3.16 对参数线性规划$(\mathrm{LP}\hat{\mathrm{C}}_\theta)$，设其对某个 θ_0有最优解，则

(i) *存在某个包含 θ_0的闭区间$[\underline{\theta},\overline{\theta}]$：$(\mathrm{LP}\hat{\mathrm{C}}_\theta)$对 $\theta\in[\underline{\theta},\overline{\theta}]$有最优解，对 $\theta<\underline{\theta}$及 $\theta>\overline{\theta}$无最优解；*

(ii) 设 $\theta=\theta_0$时的最优基本可行解为 x^0，则存在$\underline{\theta}_0$和$\overline{\theta}^0$，使得：当 $\theta\in[\underline{\theta}_0,\overline{\theta}_0]$时 x^0都是最优解，而当 $\theta\notin[\underline{\theta}_0,\overline{\theta}_0]$时 x^0不是最优解(可能$\underline{\theta}_0=-\infty$，或$\overline{\theta}_0=+\infty$)；

(iii) 若$\underline{\theta}_0<\bar{\theta}$,则存在一个基本可行解 x^1,它对 $\theta=\bar{\theta}_0$是最优解,而对 $\theta<\bar{\theta}_0$不是最优解;若$\underline{\theta}_0>\underline{\theta}$,则存在一个基本可行解 x^2,它对 $\theta=\underline{\theta}_0$是最优解,而对 $\theta>\underline{\theta}_0$不是最优解;

(iv) $(\mathrm{LP}\hat{\mathrm{C}}_\theta)$的最优值 $v(\theta)$在$[\underline{\theta},\bar{\theta}]$上是连续的、分段线性的凹函数.

给定了定理中的 θ_0,则其中的$\underline{\theta}_0,\bar{\theta}^0,x^1,x^2,\underline{\theta},\bar{\theta}$都可以用单纯形法求得(参见定理3.15以下的一段话中给出的方法). 此时,不等式(3.30)都是线性不等式,求解很容易.

2. 等式约束常数项含参数的线性规划

考虑问题

$$(\mathrm{LPB}_\theta)\qquad \begin{aligned} \min\ & c^T x \\ \text{s.t.}\ & Ax=b+\theta\hat{b} \\ & x\geqslant 0 \end{aligned}$$

其中 $\hat{b}\in R^m$ 为给定. 它的对偶问题为

$$(\mathrm{LDB}_\theta)\qquad \begin{aligned} \max\ & (b+\theta\hat{b})^T\lambda \\ \text{s.t.}\ & A^T\lambda\leqslant c \end{aligned}$$

定义新的变量 $u,v,w\in R_+^m$,以及 $\lambda=u-v$,则上规划可化为等价的

$$(\mathrm{LD}\hat{\mathrm{B}}_\theta)\qquad \begin{aligned} \min\ & -(b+\theta\hat{b})^T(u-v) \\ \text{s.t.}\ & A^T(u-v)+w=c \\ & u,v,w\geqslant 0 \end{aligned}$$

由下面的关系

(LPB_θ) 有最优解 $\Leftrightarrow(\mathrm{LDB}_\theta)$ 有最优解 $\Leftrightarrow(\mathrm{LD}\hat{\mathrm{B}}_\theta)$ 有最优解

可知,对(LPB_θ)的研究可转化对$(\mathrm{LD}\hat{\mathrm{B}}_\theta)$的研究,而$(\mathrm{LD}\hat{\mathrm{B}}_\theta)$与$(\mathrm{LP}\hat{\mathrm{C}}_\theta)$的形式是相同的. 于是可得到与定理3.16之(i) ~(iii)相类似的结果. 又由于 (LPB_θ)的最优值 $\hat{v}(\theta)$与$(\mathrm{LD}\hat{\mathrm{B}}_\theta)$的最优值的相反数相等. 类似于(3.28),设$(\mathrm{LD}\hat{\mathrm{B}}_\theta)$的基本可行解集合为

$$\left\{\begin{bmatrix}u^1\\v^1\\w^1\end{bmatrix},\begin{bmatrix}u^2\\v^2\\w^2\end{bmatrix},\cdots,\begin{bmatrix}u^p\\v^p\\w^p\end{bmatrix}\right\}\qquad(3.33)$$

再记

$$\lambda^i = u^i - v^i,\ i = 1,\cdots,p \tag{3.34}$$

于是就有

$$\hat{v}(\theta) = \max_{1\leqslant i\leqslant p}\{(b+\theta\hat{b})^T\lambda^i\} = \max_{1\leqslant i\leqslant p}\{\hat{l}_i(\theta)\} \tag{3.35}$$

其中

$$\hat{l}_i(\theta) = (b+\theta\hat{b})^T\lambda^i = (\hat{b}^T\lambda^i)\theta + (b^T\lambda^i) \tag{3.36}$$

上式可以看成是(3.32)的对偶形式．容易从图像中看出，由(3.35)给出的 $\hat{v}(\theta)$是 R 上的连续的、分段线性的(下)凸函数．

定理3.17 对参数线性规划(LPB_θ)，设其对某个 θ_0有最优解，则

(i) 存在某个包含 θ_0的闭区间$[\underline{\theta},\overline{\theta}]$：($\mathrm{LPB}_\theta$)对 $\theta\in[\underline{\theta},\overline{\theta}]$有最优解，而对 $\theta\notin[\underline{\theta},\overline{\theta}]$无最优解；

(ii) 设 $\theta=\theta_0$时的最优基为 B_0，则存在$\underline{\theta}_0$和$\overline{\theta}_0$，使得：当 $\theta\in[\underline{\theta}_0,\overline{\theta}_0]$时 B_0都是最优基，而当 $\theta\notin[\underline{\theta}_0,\overline{\theta}_0]$时 B_0不是最优基(可能$\underline{\theta}_0=-\infty$，或$\overline{\theta}_0=+\infty$)；

(iii) 若$\overline{\theta}_0<\overline{\theta}$，则存在一个对偶可行基 B_1，它对 $\theta=\overline{\theta}_0$是最优基，而对 $\theta<\overline{\theta}_0$不是最优基；若$\underline{\theta}_0>\underline{\theta}$，则存在一个对偶可行基 B_2，它对 $\theta=\underline{\theta}_0$是最优基，而对 $\theta>\underline{\theta}_0$不是最优基；

(iv) (LPB_θ)的最优值 $\hat{v}(\theta)$在$[\underline{\theta},\overline{\theta}]$上是连续的，分段线性的凸函数．

为了求得定理中的$\underline{\theta}_0,\overline{\theta}_0,B_1,B_2,\underline{\theta},\overline{\theta}$，设($\mathrm{LPB}_{\theta_0}$)已有了一个最优基 B_0，写出(LPB_θ)关于基 B_0的单纯形表：

$$\begin{bmatrix} B_0^{-1}A & B_0^{-1}(b+\theta\hat{b}) \\ c^T - c_{B_0}^T B_0^{-1}A & -c_{B_0}^T B_0^{-1}(b+\theta\hat{b}) \end{bmatrix} \tag{3.37}$$

表中已有 $c^T-c_{B_0}^T B_0^{-1}A\geqslant 0$．为保持可行性，应考虑线性不等式组：

$$B_0^{-1}(b+\theta\hat{b})\geqslant 0 \tag{3.38}$$

利用对偶单纯形法，与(3.29)以下的一段话作“对偶的”分析可得所需的方法．详述略．

最后我们指出，如果线性规划(LP)是非退化的，于是在表(3.37)中总有 $B_0^{-1}b>0$．从而对充分小的 θ 亦有

$$B_0^{-1}(b+\theta\hat{b})>0 \tag{3.39}$$

设 B_0 为(LP)的最优基，于是对充分小的 θ，(LPB_θ)的最优值就是 $c_{B_0}^T B_0^{-1}(b+\theta\hat{b})$. 记 $\lambda^0=(c_{B_0}^T B^{-1})^T$，它是(LP)的对偶问题的最优解. 于是对充分小的 θ，(LPB_θ)的最优值为

$$b^T\lambda^0+(\theta\hat{b})^T\lambda^0 \tag{3.40}$$

由于此时 $b^T\lambda^0$ 也是(LP)的最优值，(3.40)表明对非退化的(LP)，b 经过小扰动 $\theta\hat{b}$ 后成为 $b+\theta\hat{b}$，最优值的变化就是 $(\theta\hat{b})^T\lambda^0$，而(LP)的对偶的最优解 λ^0 就是相应的变化率.

习　题

1. 验证对称对偶(3.1)式的对合性，即对偶的对偶是原问题.

2. 对于对称对偶(3.1)式，

(i) 写出并证明其松紧定理；

(ii) 设已知例1.1有最优解(1000，2500)，利用松紧定理求解其对偶问题(3.3).

3. 构造一个原始问题，它本身及其对偶均无可行解.

4. Farkas 引理：对所有满足 $Ax\geqslant 0$ 的 x 都有 $b^Tx\geqslant 0$，其充要条件是存在 $y\geqslant 0$ 使得 $A^Ty=b$.

证明 Farkas 引理与对偶定理是等价的.

5. 双线性 Stackelberg 问题如下：

(BSP)
$$\max_{x\in X} F(x)$$

其中 X 是 R^n 中的非空子集，且

$$F(x)=\min_{y\in Y_0(x)} x^TAy+a^Tx+c^Ty$$

$$Y_0(x)=\{y\in Y: b^Ty=\max_{w\in Y} b^Tw\}$$

$$Y=\{y\in R^m: Gy=g, y\geqslant 0\}$$

A，G 都是给定的矩阵，a，b，c，g 都是给定的向量，它们维数恰使以上的表达式有意义.

利用对偶理论将上面的问题(BSP)化为

(BSD)
$$\begin{aligned}\max_{x,z,u,\lambda}\ & b^Tz-g^Tu+a^Tx\\ \text{s.t.}\ & A^Tx+G^Tu-b\lambda\geqslant -c\\ & Gz=\lambda g,\ z\geqslant 0,\lambda\geqslant 0\\ & x\in X\end{aligned}$$

写出这两个规划的最优值和最优解之间的关系. 注意,若 X 由线性等式或线性不等式所描述,则(BSD)已是线性规划,而(BSP)却是比较复杂的问题.

6. (i) 用单纯形法求解

$$\begin{aligned}\min\ & 2x_1+3x_2+2x_3+2x_4\\ \text{s.t.}\ & x_1+2x_2+x_3+2x_4=3\\ & x_1+x_2+2x_3+4x_4=5\\ & x_i\geqslant 0,\ i=1,2,3,4\end{aligned}$$

(ii) 将等式约束右端分别改为8和7,再利用对偶单纯形法解这个新问题.

7. 证明由(3.22)定义的 $\bar{B}$ 是那里问题的对偶可行基.

8. 假设线性规划(LP)是非退化的. 不用对偶理论而用直接在单纯形表(3.29)上作取主运算的方法证明定理3.16.

9. 作一个由(3.28)所定义的 $v(\theta)$ 的一个图像,其中 $c(\theta)$ 取为一般的凹函数;再进一步理解定理3.15下面有关求解 $\bar{\theta}$ 的那段话.

10. 对所有的 θ 求解线性规划

$$\begin{aligned}\max\ & (1+\theta)x_1-(1-\theta)x_2\\ \text{s.t.}\ & x_1+x_2+x_3=3\\ & x_1-2x_2+x_4=1\\ & -2x_1+x_2+x_5=2\\ & x_j\geqslant 0,\ j=1,2,3,4,5\end{aligned}$$

11. 研究原始线性规划问题

$$\begin{aligned}\min\ & c^Tx\\ \text{s.t.}\ & Ax=b\\ & x\geqslant 0\end{aligned}$$

假定这个问题与其对偶都有可行解. 令 λ 是对偶问题的一个最优解.

(i) 若用 $\mu\neq 0$ 乘以原问题的第 k 个方程,试确定所得新问题的对偶最优解 w;

(ii) 将原始问题的第 k 个方程的 μ 倍加到第 r 个方程上,新问题的对偶最优解是什么?

(iii) 将原始问题中 A 的第 k 行的 μ 倍加到 c 上,所得新问题的对偶最优解又是什么?

12. 用第一章习题5的结果直接证明引理3.13.

第四章 运输问题

运输问题的数学模型如下：

已知 $a_i>0, i=1,\cdots,m; b_j>0, j=1,\cdots,n$，满足

$$\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j \tag{4.1}$$

及 $m\times n$ 阵 $(c_{ij}), c_{ij}\geqslant 0$.

求 $m\times n$ 阵 (x_{ij})，其为下问题的解：

$$\min \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}x_{ij} \tag{4.2}$$

$$\text{s.t.} \sum_{j=1}^{n} x_{ij} = a_i,\ i=1,\cdots,m \tag{4.3}$$

$$\sum_{i=1}^{m} x_{ij} = b_j,\ j=1,\cdots,n \tag{4.4}$$

$$x_{ij}\geqslant 0,\ i=1,\cdots,m;\ j=1,\cdots,n$$

若记(4.1)的值为 Q，则取

$$x_{ij} = a_i b_j/Q$$

易知这样的(x_{ij})便是运输问题的一个可行解. 又易知可行解集是有界的闭集. 故运输问题必有解.

运输问题是形成最早的线性规划问题. 由于它的结构的特殊性，到目前为止已有很多有效的求解方法. 我们这里只介绍标号法，或称网络流法.

我们已经知道，运输问题的对偶问题是

$$\max \sum_{i=1}^{m} a_i u_i + \sum_{j=1}^{n} b_j v_j$$

$$\text{s.t.}\ u_i + v_j \leqslant c_{ij},\ i=1,\cdots,m; j=1,\cdots,n$$

将松紧定理(定理3.8)用于上两规划，我们得到

定理4.1(运输问题的判别定理) (x_{ij}^0)是运输问题的最优解的充要条件是存在(u_i^0,v_j^0)：

$$\sum_{j=1}^{n} x_{ij}^0 = a_i,\ i = 1,\cdots,m \tag{4.5}$$

$$\sum_{i=1}^{m} x_{ij}^0 = b_j,\ j = 1,\cdots,n \tag{4.6}$$

$$u_i^0 + v_j^0 \leqslant c_{ij},\ i = 1,\cdots,m;\ j = 1,\cdots,n \tag{4.7}$$

并且

$$\text{当 } u_i^0 + v_j^0 < c_{ij} \text{ 时},x_{ij}^0 = 0 \tag{4.8}$$

$$\text{当 } u_i^0 + v_j^0 = c_{ij} \text{ 时},x_{ij}^0 \geqslant 0 \tag{4.9}$$

现在介绍方法.其基本思想是从一组(u_i,v_j)到另一组(u_i,v_j)，使得它们总满足条件(4.7)；并且当$u_i+v_j<c_{ij}$时规定$x_{ij}=0$，这样条件(4.8)也满足；与此同时，对其余的x_{ij}在非负的条件下(即满足(4.9)式)进行适当调整，最后使x_{ij}满足(4.5)、(4.6).这样便得运输问题的最优解.因此这种解法实质上是对偶方法.

运输问题的标号法

第一步 求(4.7)的一个初始解，例如可取

$$u_i^0 = \min_{1\leqslant j\leqslant n} c_{ij},\ i = 1,\cdots,m$$

$$v_j^0 = \min_{1\leqslant i\leqslant m} (c_{ij} - u_i^0),\ j = 1,\cdots,n$$

注 $u_i^0+v_j^0\leqslant u_i^0+(c_{ij}-u_i^0)=c_{ij}$

第二步 利用(u_i^0,v_j^0)形成辅助问题

$$\min \sum_{i=1}^{m} y_i + \sum_{j=1}^{n} z_j\left(= 2\sum_{i=1}^{m} y_i\right)$$

$$\text{s.t.}\ \sum_{j=1}^{n} x_{ij} + y_i = a_i,\ i = 1,\cdots,m$$

$$\sum_{i=1}^{m} x_{ij} + z_j = b_j,\ j = 1,\cdots,n$$

$$x_{ij} \geqslant 0,\ y_i \geqslant 0,\ z_j \geqslant 0$$

$$x_{ij} = 0,\ \text{当 } u_i^0 + v_j^0 < c_{ij}$$

为了描述求解辅助问题的算法，我们把问题表为下表

表 一

$i \backslash j$	v_1^0	v_2^0		v_j^0		v_n^0	
u_1^0							a_1
u_2^0							a_2
u_i^0				i,j			a_i
u_m^0							a_m
	b_1	b_2		b_j		b_n	

表中每一小方格(i,j)对应于变量x_{ij}. 称使$u_i^0+v_j^0=c_{ij}$的方格为活方格；使$u_i^0+v_j^0<c_{ij}$的方格为死方格. 那么辅助问题的求解可以描述为：在各活方格中填入非负数x_{ij}，要求各行之和不超过a_i；各列之和不超过b_j，且使填入的各数之和为最大.

求出使$u_i^0+v_j^0=c_{ij}$的所有(i,j)，定出活方格；

第三步　求解辅助问题

1° 给定辅助问题的一个初始可行解(x_{ij}^0,y_i^0,z_j^0)，若所有的$y_i^0=0$，则(x_{ij}^0)已满足(4.5)、(4.6)，即为最优解，算法结束；否则，将所有的$y_i^0>0$的行都标以$(+)$，然后执行

2° 检查每一新被标上号的行，例如第i行，将该行中所有活方格所对应的尚未标上号的列，都标以$(+i)$；

3° 检查每一新被标上号的列，例如第j列，若这时$z_j^0>0$则转到4°；否则，将该j列中所有使$x_{ij}>0$但尚未标上号的行都标以$(-j)$，转回2°；

若算法终止于2°或3°，则转入第四步.

4° 设列j_t使$z_{j_t}^0>0$，且得到了标号$(+i_t)$. 再设i_t行的标号为$(-j_{t-1})$，j_{t-1}列的标号为$(+i_{t-1})$，…，依次追踪返回，可得如下的活方格序列：

$$
\begin{array}{llllll}
(i_0, & j_0) & (i_1, & j_0) & (i_1, & j_1) \quad (i_2, \; j_1) \quad \cdots \\
(+)(+i_0) & & (-j_0) & & (+i_1) & (-j_1) \\
(i_{t-1}, & j_{t-2}) & (i_{t-1}, & j_{t-1}) & (i_t, & j_{t-1}) \quad (i_t, \; j_t) \\
(-j_{t-2}) & & & (+i_{t-1}) & (-j_{t-1}) & \quad\quad (+i_t)
\end{array}
\tag{4.10}
$$

其中按标号规则有

$$y_{i_0}^0>0, \; z_{j_t}^0>0$$

$$x_{i_{k+1}, j_k}^0>0, \; k=0,1,\cdots,t-1$$

在标号列(4.10)上，对解(x_{ij}^0, y_i^0, z_j^0)作调整如下：记

$$\theta_1 = \min_{0 \leqslant k \leqslant t-1} x_{i_{k+1}, j_k}^0$$

$$\theta = \min\{\theta_1, y_{i_0}^0, z_{j_t}^0\}$$

取

$$\bar{y}_{i_0}=y_{i_0}^0-\theta$$

$$\bar{x}_{i_k, j_k}=x_{i_k, j_k}^0+\theta, \; k=0,1,\cdots,t$$

$$\bar{x}_{i_{k+1}, j_k}=x_{i_{k+1}, j_k}^0-\theta, \; k=0,1,\cdots,t-1$$

$$\bar{z}_{j_t}=z_{j_t}^0-\theta$$

而其余的 $\bar{x}_{ij}=x_{ij}^0, \bar{y}_i=y_i^0, \bar{z}_j=z_j^0$

以$(\bar{x}_{ij}, \bar{y}_i, \bar{z}_j)$代替$(x_{ij}^0, y_i^0, z_j^0)$，转到1°；

注1 根据 θ 的取法，容易验证$(\bar{x}_{ij}, \bar{y}_i, \bar{z}_j)$仍是辅助问题的可行解，其目标函数值较前减少了正数$2\theta$.

注2 假如终止于2°或3°，即既未发现已标号的列 j，使 $z_j^0>0$；也无新获标号的行或列. 这时已获辅助问题的最优解. 兹证如下：将行和列重新排成下表的形式

表 二

	J(已标号的列)	$\bar{J}$(未标号的列)
I(已标号的行)	A	$B \quad x_{ij}^0=0$
$\bar{I}$(未标号的行)	$C \quad x_{ij}^0=0$	D

从表中，我们可以看到下列事实：

(a) B 中没有活方格(不然的话；$\bar{J}$ 中的列就要被标号)；

(b) C 中所有的方格有 $x_{ij}^0=0$(不然的话,$\bar{I}$ 中就有行被标号);

(c) $\bar{I}$ 中的行 i 都使 $y_i^0=0$(不然的话,一开始就被标上号);

(d) J 中的列 j 都使 $z_j^0=0$(不然的话,已转到4°).

因此,对辅助问题的任何可行解(x_{ij},y_i,z_j),注意到在 B 中的 $x_{ij}=0$,有

$$\sum_i \sum_j x_{ij} \leqslant \sum_{i\in I} a_i + \sum_{j\in J} b_j = \sum_i \sum_j x_{ij}^0$$

这就证明了(x_{ij}^0,y_i^0,z_j^0)是辅助问题的最优解.

第四步　改进对偶可行解

设在第三步终止时,我们得到了表二. 因为 B 中没有活方格,故

$$\delta = \min_{i\in I, j\in \bar{J}} \{c_{ij} - u_i^0 - v_j^0\} > 0$$

取

$$\begin{cases} \bar{u}_i = u_i^0 + \delta, & i \in I \\ \bar{u}_i = u_i^0, & i \in \bar{I} \\ \bar{v}_j = v_j^0 - \delta, & j \in J \\ \bar{v}_j = v_j^0, & j \in \bar{J} \end{cases}$$

以$(\bar{u}_i,\bar{v}_j)$代替(u_i^0,v_j^0),转到第二步;

注1　按取法易检

$$\bar{u}_i + \bar{v}_j = \begin{cases} u_i^0 + v_j^0 \leqslant c_{ij}, & i \in I, j \in J \\ u_i^0 + v_j^0 + \delta \leqslant c_{ij}, & i \in I, j \in \bar{J} \\ u_i^0 + v_j^0 - \delta < c_{ij}, & i \in \bar{I}, j \in J \\ u_i^0 + v_j^0 \leqslant c_{ij}, & i \in \bar{I}, j \in \bar{J} \end{cases} \tag{4.11}$$

且

$$\begin{aligned} &\Big(\sum_i a_i\bar{u}_i + \sum_j b_j\bar{v}_j\Big) - \Big(\sum_i a_iu_i^0 + \sum_j b_jv_j^0\Big) \\ &= \delta\Big(\sum_{i\in I} a_i - \sum_{j\in J} b_j\Big) = \delta\Big(\sum_{i\in I} a_i - \sum_{i\in I}\sum_{j\in J} x_{ij}^0\Big) > 0 \end{aligned}$$

故改进了对偶可行解.

注2 在 a_i, b_j, c_{ij} 是整数的假定下，算法可在有限步内求得运输问题(包含其对偶问题)的最优整数解，只要我们所给的初始可行解 (x_{ij}^0, y_i^0, z_j^0) 是整数时，这是因为此时第三步中的 θ 和第四步中的 δ 都是正整数，故调整后的解也必然是整数解，且目标函数值每次都改进一正整数.

注3 在解新辅助问题时，可利用前辅助问题的最优解作为初始可行解. 对照(4.11)和表二可知，新旧辅助问题在 A、D 中的活方格没有变动，而且 B 和 C 中的 $x_{ij}^0=0$，所以本注之方法可行.

例4.2 一个运输问题可用下表给出：

单位费用 发点 \ 收点	B_1	B_2	B_3	B_4	产量
A_1	13	11	15	20	2
A_2	17	14	12	13	6
A_3	18	18	15	12	7
需要量	3	3	4	5	

1)取 $u_1^0=11, u_2^0=u_3^0=12, v_1^0=2, v_2^0=v_3^0=v_4^0=0$

2)解辅助问题

u_i^0 \ v_j^0	2	0	0	0	a_i
11	0	0	4	9	2
12	3	2	0	1	6
12	4	6	3	0	7
b_j	3	3	4	5	

方格内的数为

$$c'_{ij} = c_{ij} - u_i^0 - v_j^0$$

因此使 $c'_{ij}=0$ 的为活方格

⟶

			(+2)	(+3)	
	0	2			2
(+)			4		6
(+)				5	7
	3	3	4	5	

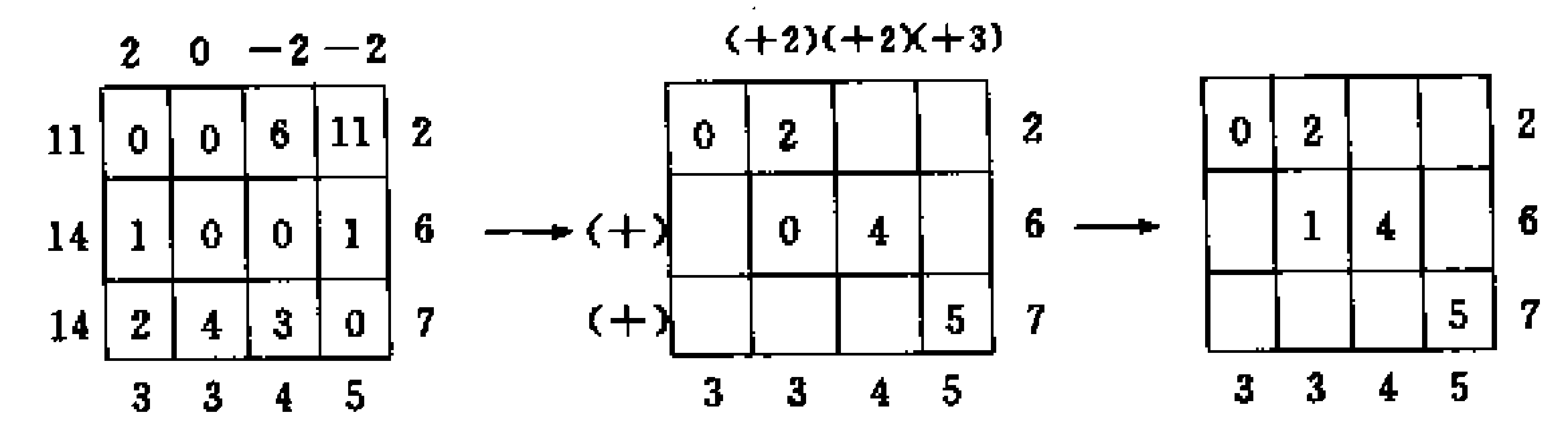
在活方格内填数、标号，标号终止于第三步的3°；故得辅助问题的最优解. 取 $\delta=\min\{3,2,4,6\}=2$.

3)改进对偶解

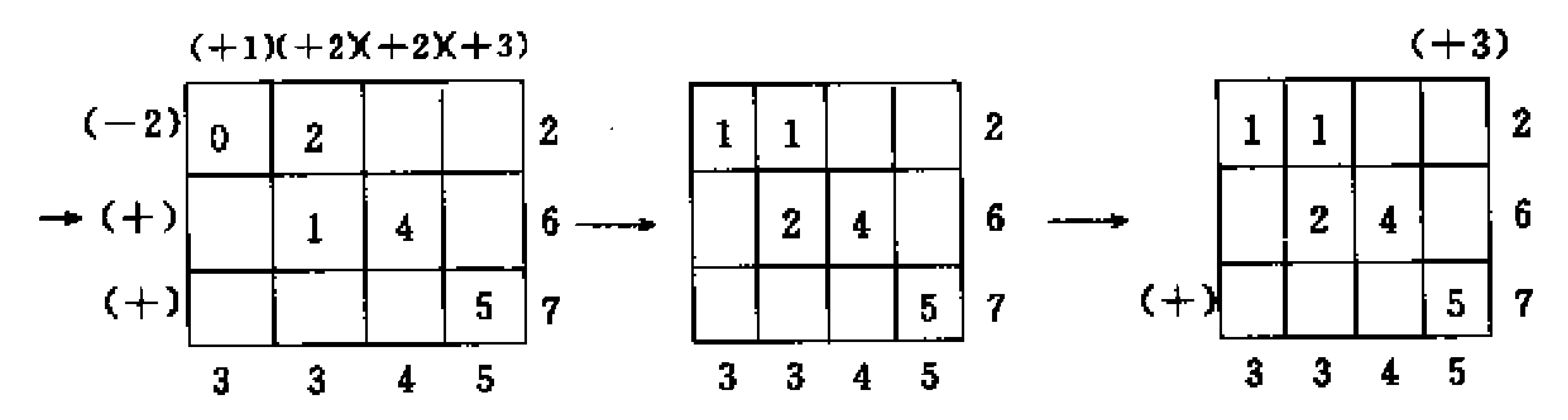
$$u_1 = 11, u_2 = u_3 = 14,$$

$$v_1 = 2, v_2 = 0, v_3 = v_4 = -2$$

4)解辅助问题



将原最优解填入，在余下的活方格内加零;标号，得标号序{(2,2)}，取$\theta=\min\{6-4,3-2\}=1$,调整得



标号,得标号序

{(2 , 2) , (1 , 2) , (1 , 1)}
(+)(+2) (−2) (+1)

取 $\theta=\min\{x_{12}^{0},6-4-1,3-0\}=1$,

调整得

标号,知已得辅助问题之最优解.取

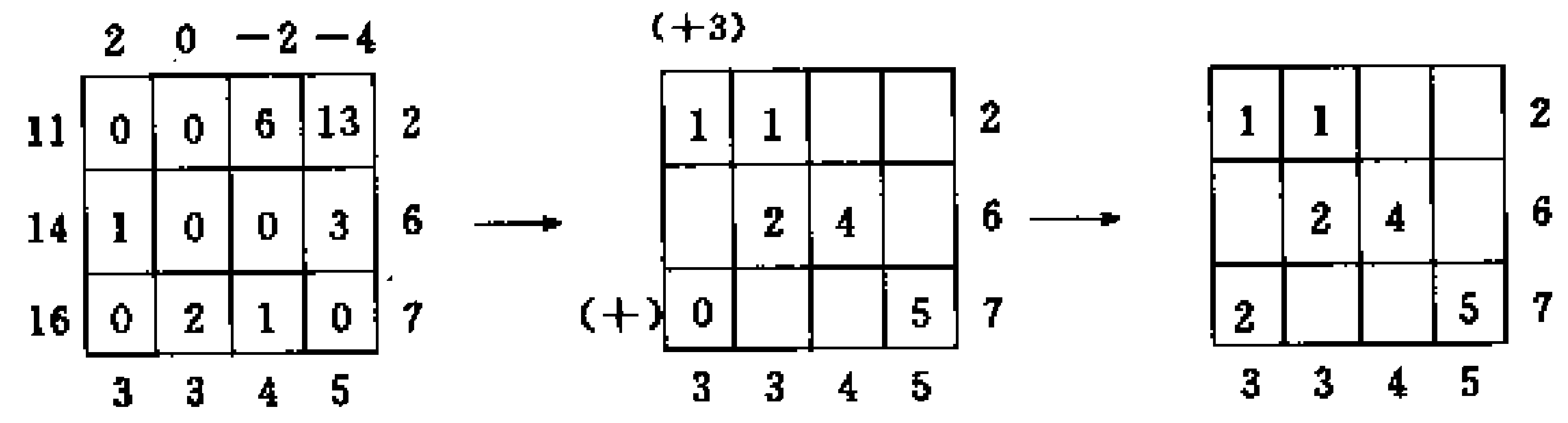
$\delta=\min\{2,4,3\}=2$

5)改进对偶解

$$u_1=11,u_2=14,u_3=16,$$

$$v_1=2,v_2=0,v_3=-2,v_4=-4$$

6)解辅助问题



标号,得标号序列{(3,1)}，取 $\theta=\min\{7-5, 3-1\}=2$,调整后得

无任何行可标号，故得运输问题的解.(对偶最优解就是5)中给出的改进对偶解.)

习　题

1. 求解下述运输问题

c_{ij}	B_1	B_2	B_3	B_4	a_i
A_1	8	7	3	2	1
A_2	4	7	5	1	9
A_3	2	4	9	6	4
b_j	3	2	4	5	

2. 证明在运输问题中线性等式约束是线性相关的. 从而再证明其对偶问题的解不唯一.

第二部分　非线性规划

第五章　非线性规划问题

§5.1　引言及基本概念

非线性规划的一般模型为

(P) $$\min_{x\in S} f(x)$$

其中 S 为 R^n 的子集，$f(x)$ 定义在 S 上或 R^n 上.

当 $S=R^n$ 时，对应的规划(P)称为无约束问题；当 S 为 R^n 的子集时，对应的(P)称为约束问题. 问题 $\max_{x\in S} f(x)$ 可化为等价的 $\min_{x\in S} -f(x)$，故仅需考虑极小化问题. 称 S 为(P)的可行集，可行集中的点 $\bar{x}$ 称为可行点，或说点 $\bar{x}$ 对(P)可行，称 $f(x)$ 为(P)的目标函数. 使 f 在 S 上取到极小的点 x^* 称为(P)的最优解，或称为(P)的解，对应的目标函数值称为问题(P)的最优值. 极小点的确切定义如下：

定义5.1　设 S 为 R^n 中的非空集，$f:S\to R$. 对 $x^*\in S$，若存在 $\varepsilon>0$ 使得

$$f(x)\geqslant f(x^*),\forall\, x\in S\cap O(x^*,\varepsilon) \tag{5.1}$$

其中 $O(x^*,\varepsilon)$ 为 R^n 中的以 x^* 为球心 ε 为半径的开球，即

$$O(x^*,\varepsilon)=\{x\in R^n:\|x-x^*\|<\varepsilon\}$$

则说 x^* 是 f 在 S 上的一个局部极小点；若

$$f(x)\geqslant f(x^*),\forall\, x\in S \tag{5.2}$$

则说 x^* 是 f 在 S 上的一个全局极小点. 又若(5.1)(或(5.2))对 $x\neq x^*$ 严格成立，则说 x^* 是 f 在 S 上的一个严格局部(或严格全局)极小点.

特别地，我们称问题

(NP)
$$\begin{aligned}&\min f(x)\\&\text{s.t. } h_i(x)=0,\ i=1,\cdots,m\\&\qquad g_j(x)\leqslant 0,\ j=1,\cdots,p\end{aligned}$$

为标准非线性规划. 其中 f,h_i,g_j 都是 R^n 上的实值函数，"s.t."表示"受限制于"，取自英文"subject to"的缩写. 称 $h_i(x)=0$为等式约束；称 $g_j(x)\leqslant 0$为不等式约束. 若采用向量的形式，记

$$h(x)=(h_1(x),h_2(x),\cdots,h_m(x))^T \tag{5.3}$$

$$g(x)=(g_1(x),g_2(x),\cdots,g_p(x))^T \tag{5.4}$$

则有紧凑的形式

(NP)
$$\begin{aligned}&\min f(x)\\&\text{s.t. } h(x)=0\\&\qquad g(x)\leqslant 0\end{aligned}$$

在 $f,h_i,i=1,\cdots,m,g_j,j=1,\cdots,p$ 都为线性函数时，对应的问题称为线性规划. 线性规划的一种标准型为

(LP)
$$\begin{aligned}&\min c^Tx\\&\text{s.t. } Ax=b\\&\qquad x\geqslant 0\end{aligned}$$

其中 $A\in R^{m\times n},c\in R^n,b\in R^m$ 为固定，$x\in R^n$ 为变量. 关于线性规划的一般理论和算法，可参见[4]、[11]、[24]和[25].

如果在(NP)中，f 和 g_j 都是凸函数且 h_i 皆为线性函数，则称这一类问题为标准凸规划. 这种规划特别便于分析，它具有某些有趣的性质，在理论和实际应用中都吸引着人们，有关凸函数的许多性质都可以在[17]中找到.

第五章至第九章主要讨论非线性规划的理论问题. 这包括给出最优解的判别准则，即最优性条件(充分条件或必要条件)；研究非线性规划问题在最优解处的性质，对偶和鞍点，等等. 第十章至第十三章讨论如何求解一个非线性规划问题，这方面的内容也称为最优化方法. 如不作特别说明，求解一个非线性规划问题的涵义是求一个最优解向量 x^*，而不是求可能存在的所有最优解. 应该

指出，本部分的某些结果也适用于抽象空间中的分析，有兴趣的读者不妨试着作一些推导. [1]、[2]和[12]都是非线性规划方面的较好的参考书，[23]对无约束问题的求解方法有较全面的论述.

一般若不作说明，带有下标的字母（如 h_i, x_i, x_{ij}）表示数量，带有上标的字母（如 x^k, d^k）代表向量. 设 z 为向量，$z \geqslant 0$ 表 z 的每个分量非负，$z > 0$ 表 z 的分量皆正，$z = 0$ 表 z 的分量皆为零，等等.

设 f 在 R^n 上连续可微（记为 $f \in C^1$），记 $x = (x_1, x_2, \cdots, x_n)^T$，再记

$$\nabla f(x) = \left(\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \cdots, \frac{\partial f(x)}{\partial x_n} \right)^T$$

称 $\nabla f(x)$ 为 f 在 x 处的梯度. 再设 f 在 R^n 上二次连续可微（记为 $f \in C^2$），记

$$\nabla^2 f(x) = \begin{pmatrix} \frac{\partial^2 f(x)}{\partial x_1 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \cdots & \frac{\partial^2 f(x)}{\partial x_n \partial x_1} \\ \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \frac{\partial^2 f(x)}{\partial x_2 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} \\ \cdots & \cdots & & \\ \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} & \cdots & \frac{\partial^2 f(x)}{\partial x_n \partial x_n} \end{pmatrix}$$

称 $\nabla^2 f(x)$ 为 $f(x)$ 在 x 处的二阶导数阵，或 Hesse 阵.

§5.2 几个实例

非线性规划在工程、军事、经济学、企业管理、物理学等方面都有重要应用. 这里仅举数例说明.

例5.2（Sylvester 问题） 设平面上有 m 个点，找覆盖这 m 个点的最小圆盘.

设这 m 个点为 $p_i, i = 1, \cdots, m$. 平面上任一点 x 到各点距离之最大者为 $f(x) = \max_{1 \leqslant i \leqslant m} \{\|x - p_i\|\}$. 以 x 为圆心，$f(x)$ 为半径的圆盘必覆盖这 m 个点，于是问题转化为求解下面的非线性规划：

$$\min_{x\in R^2}\max_{1\leqslant i\leqslant m}\{\|x-p_i\|\} \tag{5.5}$$

它是一个无约束问题，其目标函数 $f(x)$ 在某些点不可微. 请读者证明，上问题可化为求解如下的可微约束优化问题：

$$\begin{cases}\min\limits_{r,x} r \\ \text{s.t. } r\geqslant\|x-p_i\|^2,\ i=1,\cdots,m \\ \qquad r\in R\end{cases} \tag{5.6}$$

例5.3（场址问题） 设有 n 个商店，其位置和对货物的需求都是已知的；其货物由 m 个仓库提供，仓库的容量也是已知的. 决定这 m 个仓库建于何处，使得由仓库提供各商店货物中的运量与路程之积之总和为最小.

设 (x_i,y_i) 为仓库的设址，$i=1,\cdots,m$，

c_i 为仓库的库容量，$i=1,\cdots,m$，

(a_j,b_j) 为商店的位置，$j=1,\cdots,n$，

r_j 为商店对货物的需求量，$j=1,\cdots,n$，

d_{ij} 为从第 i 个仓库到第 j 个商店的路程，$\forall\ i,j$，

w_{ij} 为从第 i 个仓库到第 j 个商店的运量，$\forall\ i,j$.

于是场址问题的数学模型为

$$\begin{cases}\min\ \sum\limits_{i=1}^{m}\sum\limits_{j=1}^{n}d_{ij}w_{ij} \\ \text{s.t. } \sum\limits_{j=1}^{n}w_{ij}\leqslant c_i,\ i=1,\cdots,m \\ \qquad \sum\limits_{i=1}^{m}w_{ij}=r_j,\ j=1,\cdots,n \\ \qquad w_{ij}\geqslant 0,\ i=1,\cdots,m,\ j=1,\cdots,n\end{cases} \tag{5.7}$$

注意，在问题(5.7)中 d_{ij} 和 w_{ij} 都是变量，所以它是约束为线性的非线性规划，其中的路程 d_{ij} 可各取为下两者之一：

$$d_{ij}=|x_i-a_j|+|y_i-b_j|$$

$$d_{ij}=[(x_i-a_j)^2+(y_i-b_j)^2]^{1/2}$$

又，若仓库的地址是预先固定的，则此时的 d_{ij} 都是常量，对应的问

题(5.7)便成了线性规划.这种规划作为运输问题,是最早形成的线性规划.

例5.4(供水网络中的最优设计) 设有一个供水网络,它由某个连通图给出,此图有 n 个结点,t 条弧,m 条独立的回路.假定供水只设在一个结点处,再假设在每条弧中的流量为 q_i,压强差为 p_i.我们就有下面的数学模型:

$$
\begin{cases}
\min \sum_{i=1}^{t} c_i p_i^{\alpha} q_i^{\beta} & (5.8a) \\
\text{s.t.} \sum_{i\in I(k)} q_i - \sum_{j\in O(k)} q_j = b_k, k = 1,\cdots,n & (5.8b) \\
\sum_{i\in L(l)} p_i = 0, l = 1,\cdots,m & (5.8c) \\
\sum_{i\in r(k)} p_i \leqslant h_0 - h_{k,\min}, r(k) \in R(k), k = 1,\cdots,n. & (5.8d) \\
q_{i,\min} \leqslant q_i \leqslant q_{i,\max},\ i = 1,\cdots,t & (5.8e) \\
f_i(q_i) \leqslant p_i \leqslant g_i(q_i), i = 1,\cdots,t & (5.8f)
\end{cases}
$$

其中 $I(k)$ 和 $O(k)$ 分别表示在第 k 个结点处进弧和出弧的集合;$L(l)$ 表示在回路 l 中所含的弧的集合;$r(k)$ 表示从供水结点到结点 k 的一条道路中弧的集合;$R(k)$ 表示从供水结点到结点 k 的道路的集合.于是,(5.8b)代表结点处水的消耗,(5.8c)代表回路中的平衡,(5.8e)代表弧的容量,(5.8d)和(5.8f)代表技术上的要求,$c_i,\alpha,\beta>0$ 都是预先给定的常数,p_i 和 q_i 是变量.

此问题的一般构造可参见[6].

习　题

1. 证明:若(5.5)有解 x^*,则存在 r^* 使得 (r^*,x^*) 是(5.6)的解;反之,若 (r^*,x^*) 是(5.6)的解,则 x^* 是(5.5)的解.

2. 设有数量为 x_1 的某种资源,将它投入两种生产 A 与 B.若以数量 y_1 投入生产 A,剩下的量 x_1-y_1 投入生产 B,则可得收入为

$$g(y_1) + h(x_1 - y_1)$$

其中 g 和 h 为已给的函数，且 $g(0)=h(0)=0$. 现设以 y_1 与 x_1-y_1 投入 A,B 生产后可以回收再投入生产，设它们的回收率分别为 $a,b\in[0,1]$. 因此在第一阶段生产后回收总量为

$$x_2 = ay_1 + b(x_1 - y_1)$$

将 x_2 再投入生产 A,B，然后再回收，…. 这样一共生产了 n 次. 希望选择 y_1，y_2，…，y_n 使总收入最多. 建立这个问题的数学模型.

3. 设从 a 点到 b 点建筑一条公路，路线已定. 设此公路总长为 T. 设未筑路时线路的高度函数为 $c(t)$，$t\in[0,T]$. 今要求设计公路的高度 $y(t)$，$t\in[0,T]$，使得在建筑公路时各处减少的高度（对应于挖去的土方），或增加的高度（对应于填上的土方）之总和为最少，同时要求每点处的坡度不超过 s_1，且每点处的坡度的变化率不超过 s_2. 试建立此问题的数学模型.

第六章 凸 集

读者以后将会明白，非线性规划中的某些理论的建立会以这样或那样的方式与凸集和凸函数相联系，本章第三节的内容便是很好的说明性的例子. 我们将给出与数学规划密切相关的凸集和凸函数的一些性质.

§ 6.1 凸集及其基本性质

定义6.1 设 $C\subseteq R^n$. 若对 $\forall\ x^1,x^2\in C,\forall\ \alpha\in(0,1)$ 均有

$$\alpha x^1+(1-\alpha)x^2\in C \tag{6.1}$$

则称 C 为凸集.

例如，给定 $x^*\in R^n,\varepsilon>0$，则 $O(x^*,\varepsilon)$ 是凸集；又如

$$S=\{x\in R^n:Ax=b,x\geqslant 0\}$$

也是凸集. 这些都可以直接由定义验证.

定理6.2 设 C,C_1,C_2 都是 R^n 中的凸集. 则

(i) $\alpha C=\{\alpha x:x\in C\}$ 是凸集，其中 α 是实数；

(ii) $C_1+C_2=\{x+y:x\in C_1,y\in C_2\}$ 是凸集；

(iii) $C_1\cap C_2$ 是凸集.

这个定理的证明留给读者.

设 $\lambda_j,j=1,\cdots,k$ 满足 $\lambda_j\geqslant 0$ 且 $\sum_{j=1}^k\lambda_j=1$，则称

$$x=\sum_{j=1}^k\lambda_j x^j$$

为 $\{x^1,x^2,\cdots,x^k\}$ 的一个凸组合.

定义6.3 设 $S\subseteq R^n$. S 的凸包，记为 $\mathrm{co}S$，其定义为 $x\in\mathrm{co}S$ 当且仅当存在 k 及 $x^j\in S,j=1,\cdots,k$，使得 x 是这 k 个元的某个凸组合. 显见，$\mathrm{co}S$ 是凸集.

定理6.4(Carathéodory) 设 $S\subseteq R^n$. 若 $x\in \mathrm{co}S$,则 x 可表示为 S 中至多 $n+1$ 个元的凸组合.

证 由 $x\in \mathrm{co}S$ 知,存在 $x^j\in S$, $j=1,\cdots,k$,存在 $\lambda_j\geqslant 0$, $j=1,\cdots,k$,且 $\sum_{j=1}^k \lambda_j=1$,使得 $x=\sum_{j=1}^k \lambda_j x^j$. 若 $k\leqslant n+1$ 则证明已告结束. 下设 $k>n+1$. 于是 $x^2-x^1, x^3-x^1,\cdots,x^k-x^1$ 线性相关,即存在不全为零的一组数 $\mu_2,\mu_3,\cdots,\mu_k$,使得 $\sum_{j=2}^k \mu_j(x^j-x^1)=0$. 今定义 $\mu_1=-\sum_{j=2}^k \mu_j$,于是

$$\sum_{j=1}^k \mu_j x^j = 0,\quad \sum_{j=1}^k \mu_j = 0 \tag{6.2}$$

由于上面的 μ_j 不全为零,故可设 $\mu_j>0$ 至少对某个 j 成立. 再由(6.2)之第一式得,对任意实数 α,有

$$x = \sum_{j=1}^k (\lambda_j - \alpha\mu_j)x^j \tag{6.3}$$

今取 α 为

$$\alpha = \min_{1\leqslant j\leqslant k}\{\lambda_j/\mu_j : \mu_j > 0\} = \lambda_i/\mu_i$$

容易看出

$$\lambda_j - \alpha\mu_j \geqslant 0, j = 1,\cdots,k, \text{且 } \lambda_i - \alpha\mu_i = 0 \tag{6.4}$$

并且由(6.2)之第二式推得

$$\sum_{j=1}^k (\lambda_j - \alpha\mu_j) = \sum_{j=1}^k \lambda_j - \alpha\sum_{j=1}^k \mu_j = 1 \tag{6.5}$$

(6.3)~(6.5)表明 x 可表为 S 中至多 $k-1$ 个元的凸组合. 必要时可重复以上过程,定理即获证.

Carathéodory 定理是关于维数的基本定理. 它有许多重要应用,例见习题2.

凸集 C 的闭包(记为 $\mathrm{cl}C$),内部(记为 $\mathrm{int}C$),以及其边界(记为 ∂C)与一般集合的相应的概念同. 关于凸集 C 还有一个很有用的概念叫"相对内部",记为 $\mathrm{ri}C$,可以定义如下:

$$\mathrm{ri}C = \{x \in C : \forall\, y \in C, \exists\, \alpha > 0 \text{ 使 } x - \alpha(y - x) \in C\} \tag{6.6}$$

例如,令 $C=[0,1]$,将它看成 R 中的子集则 $\mathrm{int}C=(0,1)$;若将 C

看成 R^2中介于点(0,0)和点(1,0)之间的线段,则 $\mathrm{int}C$ 为空集,但由(6.6)即推出 $\mathrm{ri}C=\{(x_1,0):0<x_1<1\}$.

已证(例如见[17]),若 C 为非空凸集则 $\mathrm{ri}C$ 非空,并且,$\forall\ x\in \mathrm{ri}C,\forall\ y\in \mathrm{cl}C$,有

$$\lambda x+(1-\lambda)y\in \mathrm{ri}C,\ \forall\ \lambda\in(0,1) \tag{6.7}$$

§6.2 凸集的分离定理

定理6.5(点与凸集的分离定理) 设 D 为 R^n 中的非空闭凸集.设 $y^0\notin D$.则存在 $c\in R^n,c\neq 0$,以及数 β,使得

$$c^Tx\geqslant\beta>c^Ty^0,\ \forall\ x\in D \tag{6.8}$$

证 任取 $z^0\in D$,作

$$Z=\{x\in D:\|x-y^0\|\leqslant\|z^0-y^0\|\}$$

则 Z 为非空有界闭凸集,而$\|x-y^0\|$是 x 的连续函数,故它在 Z 上的某点处取得最小值,设为 x^0,即

$$\|x^0-y^0\|=\min_{x\in Z}\|x-y^0\|=\min_{x\in D}\|x-y^0\| \tag{6.9}$$

上面的后一等式由 Z 的定义直接推出.由 $x^0\in D$ 及 $y^0\notin D$ 知$\|x^0-y^0\|>0$,此即

$$(x^0-y^0)^T(x^0-y^0)>0$$

上式中,令 $c=x^0-y^0$,则 $c\neq 0$;再令 $c^Tx^0=\beta$ 即得(6.8)之第二个不等式.易知,对此 c 和 β,(6.8)之第一个不等式等价于

$$(x-x_0)^T(x^0-y^0)\geqslant 0,\ \forall\ x\in D \tag{6.10}$$

任取 $x\in D$,由 D 为凸集知 $\alpha x+(1-\alpha)x^0\in D,\forall\ \alpha\in(0,1)$.故由(6.9)得

$$\|x^0-y^0\|\leqslant\|(\alpha x+(1-\alpha)x^0)-y^0\|$$

将上式两边同时平方,整理得等价的

$$\alpha^2\|x-x^0\|^2+2\alpha(x-x^0)^T(x^0-y^0)\geqslant 0$$

再以2α 遍除上式,复令 $\alpha\to 0^+$,得(6.10).证毕.

由定理6.5可证著名的 Farkas 引理:

定理6.6(Farkas)　设 $A\in R^{m\times n}$,$b\in R^m$. 则 $Ax=b,x\geqslant 0$有解的充要条件为:对任意满足 $u^TA\geqslant 0$的 u 必有 $u^Tb\geqslant 0$.

注　Farkas 引理可以转述如下:

设 $A\in R^{m\times n}$,$b\in R^m$,则下列两种情况恰有一种发生.

Ⅰ. $Ax=b,x\geqslant 0$有解 $x\in R^n$;

Ⅱ. $u^TA\geqslant 0,u^Tb<0$有解 $u\in R^m$.

类似于上面的定理统称"择一定理". 它的典型的证明步骤是:先证若Ⅰ、Ⅱ同时发生就会得到矛盾(这就证明了两者不能同时发生),这一结论一般都很容易证明;再由其中的某一个不成立去推得另一个成立(这就同时证明了若Ⅰ不发生则Ⅱ发生,以及若Ⅱ不发生则Ⅰ发生;从而证明了两者至少有一发生).

定理6.6的证明　若Ⅰ、Ⅱ同时发生,分别取Ⅰ、Ⅱ的解 $\hat{x}$ 和 $\hat{u}$,于是由

$$0\leqslant(\hat{u}^TA)\hat{x}=\hat{u}^T(A\hat{x})=\hat{u}^Tb<0$$

得矛盾. 再设Ⅰ不发生,构造集合

$$D=\{y\in R^m;y=Ax,x\in R^n,x\geqslant 0\}$$

则 $b\notin D$. 由 D 为 R^m 中的非空的闭凸集(参见本章习题6). 利用定理6.5知存在 $c\in R^m$,$c\neq 0$,使得对$\forall$ $y\in D$ 有 $c^Ty>c^Tb$,即

$$c^TAx>c^Tb,\ \forall\ x\geqslant 0$$

上式中,分别取 $x=(0,\cdots,0,x_i,0,\cdots,0)^T$,$i=1,\cdots,n$,由 x_i 可以充分大即推知 $c^TA\geqslant 0$;另一方面,令 $x=0$又得 $c^Tb<0$. 于是有 $c^TA\geqslant 0$且 $c^Tb<0$. 这表明 c 是Ⅱ的解,即Ⅱ发生. 证毕.

定理6.7(支撑超平面的存在性)　设 D 为 R^n 中的非空凸集,设 $\bar{x}\in\partial D$. 则存在 $c\in R^n$,$c\neq 0$,使得

$$c^T(x-\bar{x})\leqslant 0,\ \forall\ x\in \mathrm{cl}D \tag{6.11}$$

证　由 $\bar{x}\in\partial D$ 知存在$\{y^k\}$,$y^k\notin \mathrm{cl}D$ 且 $y^k\to\bar{x}$. 从而仿定理6.5的证法知,存在 $c^k\neq 0$使得

$$(c^k)^T(x-y^k)<0,\ \forall\ x\in D$$

上式中,可令$\|c^k\|=1$(不然,先以$\|c^k\|$遍除之,重取 c^k 为 $c^k/\|c^k\|$ 即可). 于是$\{c^k\}$有收敛的子列,不妨设 $c^k\to c$,则 $c\neq 0$. 在上式中固定

x，让 $k\to\infty$，得 $c^T(x-\bar{x})\leqslant 0$. 由 $x\in D$ 是任意的，定理证毕.

对固定的 $c\in R^n, c\neq 0, \alpha\in R$，称

$$H=\{x\in R^n: c^Tx=\alpha\}$$

为 R^n 中的一个超平面. 若在上式中取 $\alpha=c^T\bar{x}$，则由(6.11)知，定理6.7的几何意义为 D 在超平面 H 的一侧且 $\mathrm{cl}D$ 与 H 交于 $\bar{x}$. 称此 H 为 D 的支撑超平面. 请读者对定理6.5作几何解释.

推论6.8(点与凸集的弱分离定理) 设 D 为 R^n 中的非空凸集，设 $\bar{x}\notin D$，则存在 $c\in R^n, c\neq 0$，使得

$$c^T(x-\bar{x})\leqslant 0,\ \forall\ x\in \mathrm{cl}D$$

证 若 $\bar{x}\notin \mathrm{cl}D$，则在定理6.7的证明中取某个 y^k 为 $\bar{x}$，对应的 c^k 作为 c 则可；若 $\bar{x}\in \mathrm{cl}D$，由条件 $\bar{x}\notin D$ 得 $\bar{x}\in\partial D$，此时的结论由定理6.7直接得到. 证毕.

注 我们在上两个证明中避免使用"若 D 为凸集则 $\mathrm{cl}D$ 亦为凸集"这一结果.

定理6.9(两个凸集之间的弱分离定理) 设 D_1 和 D_2 都是 R^n 中的非空凸集且 $D_1\cap D_2=\varnothing$. 则存在 $c\in R^n, c\neq 0$，以及数 α，使得

$$c^Tx\geqslant\alpha\geqslant c^Ty,\ \forall\ x\in D_1, \forall\ y\in D_2 \qquad (6.12)$$

证 令

$$D=D_2-D_1=\{y-x: x\in D_1, y\in D_2\}$$

由定理6.2知 D 是凸集，由 $D_1\cap D_2=\varnothing$ 知 $0\notin D$，于是利用推论6.8知存在 $c\in R^n, c\neq 0$，使得 $c^Tz\leqslant 0, \forall\ z\in D$. 这就表明 $c^Tx\geqslant c^Ty$，$\forall x\in D_1, \forall\ y\in D_2$，此又等价于

$$\inf_{x\in D_1} c^Tx\geqslant \sup_{y\in D_2} c^Ty$$

取定一个介于上不等式左右两个量之间的某个数为 α，由此 α 结合上式即知(6.12)为真. 证毕.

定理6.10(Gordan) 设 $A\in R^{m\times n}$，则下列两种情况恰发生一种：

Ⅰ. $Ax<0$ 有解 $x\in R^n$；

Ⅱ. $u^TA=0, u\geqslant 0, u\neq 0$ 有解 $u\in R^m$.

证 若Ⅰ、Ⅱ同时成立则易得矛盾. 下设Ⅰ不成立,则集合

$$D_1 = \{z: z = Ax, x \in R^n\}$$

$$D_2 = \{z: z < 0, z \in R^m\}$$

都是 R^m 中的非空凸集且 $D_1 \cap D_2 = \varnothing$,利用定理6.9即知,存在 $c \in R^m, c \neq 0$,使得

$$c^T Ax \geqslant c^T z, \forall\, x \in R^n, \forall\, z \in R^m \text{ 且 } z < 0$$

上式中,由于 z 的每个分量都可趋于负无穷,故必有 $c \geqslant 0$;再令 $z \to 0$ 得 $c^T Ax \geqslant 0, \forall\, x \in R^n$;在此式中取 $x = -A^T c$,得 $-\|c^T A\|^2 \geqslant 0$,这就表明必有 $c^T A = 0$. 综上述,c 就是Ⅱ的解.

择一定理是数学规划理论研究的重要工具. 关于择一定理的比较系统的叙述可在文献[12]中找到.

§6.3 Farkas 引理在线性规划中的应用

本节,我们给出 Farkas 引理在线性规划中的一个重要应用. 读者可以籍此初步体会到择一定理在数学规划中的作用.

联系于线性规划的标准型

$$\text{(LP)} \qquad \begin{aligned} &\min\ c^T x \\ &\text{s.t. } Ax = b \\ &\qquad x \geqslant 0 \end{aligned}$$

有另一个线性规划

$$\text{(DP)} \qquad \begin{aligned} &\max\ \lambda^T b \\ &\text{s.t. } \lambda^T A \leqslant c^T \end{aligned}$$

称(DP)为(LP)的对偶问题(读者可参见第三章).

定理6.11 若 x^0 对(LP)可行且 λ^0 对(DP)可行,则

$$c^T x^0 \geqslant (\lambda^0)^T b$$

此类定理一般称为弱对偶定理,证明留给读者.

定理6.12 设 x^* 对(LP)可行且 λ^* 对(DP)可行. 若

$$c^T x^* = (\lambda^*)^T b \tag{6.13}$$

则 x^* 和 λ^* 分别为(LP)和(DP)的解.

证 设 x 对(LP)可行，利用定理6.11得

$$c^T x \geqslant (\lambda^*)^T b = c^T x^*$$

这表明 x^* 为(LP)的解. 类似地可证 λ^* 为(DP)的解. 证毕.

定理6.13 x^* 为(LP)的解的充要条件是，x^* 对(LP)可行且存在对(DP)可行的 λ^* 使得(6.13)成立.

证 充分性即定理6.12，下证必要性.

设 x^* 为(LP)的解. 令 $y=x-x^*$，令 D 为对角阵，其对角元记为 $d_i, i=1,\cdots,n$，定义如下：

$$d_i = \begin{cases} 0, & \text{若 } x_i^* > 0, \\ 1, & \text{若 } x_i^* = 0, \end{cases} \quad i = 1,\cdots,n$$

于是，$Dx^*=0$，且容易验证 $y=0$为问题

$$\min c^T y \quad \text{s.t. } Ay = 0,\ Dy \geqslant 0$$

的解. 从而

$$\text{凡满足} \begin{pmatrix} A \\ -A \\ D \end{pmatrix} y \geqslant 0 \text{ 的 } y \text{ 必有 } c^T y \geqslant 0$$

由 Farkas 引理知，存在非负向量 $\bar{u},\bar{v},\bar{w}$. 使得

$$\bar{u}^T A - \bar{v}^T A + \bar{w}^T D = c^T$$

在上式中令 $\lambda^* = \bar{u}-\bar{v}$，从而得到

$$(\lambda^*)^T A + \bar{w}^T D = c^T \tag{6.14}$$

再由 $\bar{w}\geqslant 0$及 D 之元皆非负，又由上式得

$$(\lambda^*)^T A \leqslant c^T \tag{6.15}$$

最后，以 x^* 与(6.14)作内积，注意到 $Dx^*=0$，得

$$(\lambda^*)^T b = (\lambda^*)^T A x^* = c^T x^* \tag{6.16}$$

(6.15)表 λ^* 对(DP)可行，(6.16)表(6.13)成立. 证毕.

上定理表明，若(LP)有解则(DP)亦有解且两规划的最优值相同. 这类结果通常称为强对偶.

上面的结果还表明，求解(LP)等价于求解以 x,λ 为变量的线性系统：

$$\begin{cases} Ax = b, x \geqslant 0 \\ A^T\lambda \leqslant c \\ (\lambda^T A - c^T)x = 0 \end{cases}$$

求解线性规划有一个单纯形法，它是由 Dantzig 在1947年给出的，并被认为是本世纪数学的一个重大成果. 但单纯形法的理论结果并不令人满意. 于是，70年代至今，对线性规划求解的研究又出现了高潮. 其中[8]、[9]和[15]都是比较具有代表性的.

Karmarkar[8]运用了一个变换将(LP)转化为

$$\text{(KLP)} \qquad \begin{aligned} \min\ & c^T x \\ \text{s.t.}\ & Ax = 0 \\ & e^T x = 1 \\ & x \geqslant 0 \end{aligned}$$

其中 $e = (1,1,\cdots,1)^T \in R^n$，$A \in R^{m\times n}$，$c \in R^n$. 以下设(KLP)的可行集不空，这等价于设(KLP)有解. 这个规划通常被称为线性规划的 Karmarkar 标准型.

现在，写出(KLP)如本节开头所述之对偶，即

$$\text{(KDP)} \qquad \begin{aligned} \max\ & w_{m+1} \\ \text{s.t.}\ & \sum_{i=1}^{m} a_{ij} w_i + w_{m+1} \leqslant c_j,\ j = 1,\cdots,n \\ & w_i \in R,\ i = 1,2,\cdots,m+1 \end{aligned}$$

其中 a_{ij}是 A 的(i,j)元，c_j 是 c 的第 j 个分量；w_i 是变量，$i=1,\cdots,m+1$. 再记 $w = (w_1,\cdots,w_m)^T$. 则此规划可改写为

$$\max_{w\in R^m} \min_{1\leqslant j\leqslant n} \left(c_j - \sum_{i=1}^{m} a_{ij} w_i\right) \tag{6.17}$$

今记(KLP)的最优值为 θ，根据以上的分析知(6.17)的最优值亦为 θ. 为书写简单起见，记

$$f_j(w) = \sum_{i=1}^{m} a_{ij} w_i - c_j,\ j = 1,\cdots,n \tag{6.18}$$

于是，对应于问题(6.17)，若记 $\phi(w) = \max\limits_{1\leqslant j\leqslant n} f_j(w)$，则

$$\min_{w\in R^m} \phi(w) \tag{6.19}$$

的最优值就是$-\theta$.

对$\forall\ \mu>0$,显然有

$$[\exp(\phi/\mu)]^{\mu} < \Big[\sum_{j=1}^{n}\exp(f_j/\mu)\Big]^{\mu} \leqslant [n\cdot\exp(\phi/\mu)]^{\mu}$$

对上式取自然对数,得

$$\phi(w) < \mu\ln\Big[\sum_{j=1}^{n}\exp(f_j(w)/\mu)\Big] \leqslant \phi(w) + \mu\ln n$$

在上式中对w在R^m上最小化,得

$$-\theta \leqslant \mu\min_{w\in R^m}\ln\Big[\sum_{j=1}^{n}\exp(f_j(w)/\mu)\Big] \leqslant -\theta + \mu\ln n$$

今将以上的推导总结如下.

定理6.14 设θ是(KLP)的最优值.则对$\forall\ \mu>0$,有

$$0 \leqslant \theta - \theta_\mu^* \leqslant \mu\ln n \tag{6.20}$$

其中

$$\theta_\mu^* = -\mu\min_{w\in R^m}\ln\Big\{\sum_{j=1}^{n}\exp\Big[\Big(\sum_{i=1}^{m}a_{ij}w_i - c_j\Big)/\mu\Big]\Big\}. \tag{6.21}$$

任取定$\varepsilon>0$,再取一个$\mu>0$满足$\mu\ln n\leqslant\varepsilon$.这定理告诉我们,对此$\mu$求单个的无约束极小化问题(6.21),可得到$\theta$的一个满足预先给定误差$\varepsilon$的一个近似$\theta_\mu^*$.

Rajasekera 和 Fang[15]推导了与此有关的一些精细的关系式,兹述于下.

设$w^* = (w_1^*,\cdots,w_m^*)^T$为(6.21)中的极小化问题的解.对(6.21)中的目标函数关于w_i在w_i^*处求导,得

$$Q^{-1}\sum_{j=1}^{n}\exp\Big[\Big(\sum_{i=1}^{m}a_{ij}w_i^* - c_j\Big)/\mu\Big]a_{ij} = 0, i = 1,\cdots,m \tag{6.22}$$

其中

$$Q = \sum_{j=1}^{n}\exp\Big[\Big(\sum_{i=1}^{m}a_{ij}w_i^* - c_j\Big)/\mu\Big] \tag{6.23}$$

再定义

$$x_j^* = Q^{-1}\exp\Big[\Big(\sum_{i=1}^{m}a_{ij}w_i^* - c_j\Big)/\mu\Big], j = 1,\cdots,n \tag{6.24}$$

容易由上面的三个式子得到

$$Ax^* = 0, e^T x^* = 1, x^* > 0$$

即 $x^* = (x_1^*, \cdots, x_n^*)^T$ 对(KLP)可行.

利用定义式(6.24),在恒等式

$$\sum_{j=1}^{n} Qx_j^* = \prod_{j=1}^{n} Q^{x_j^*}$$

中,用

$$Q x_j^* = \exp\left[\left(\sum_{i=1}^{m} a_{ij} w_i^* - c_j\right)/\mu\right]$$

代替左边和号下的第 j 项,同时用

$$Q = \exp\left[\left(\sum_{i=1}^{m} a_{ij} w_i^* - c_j\right)/\mu\right]/x_j^*$$

代替右边积号下的第 j 个 $Q, j=1,\cdots,n$. 然后将所得之等式两边取自然对数,容易得到

$$\mu \ln Q = \sum_{j=1}^{n}\left(\sum_{i=1}^{m} a_{ij} w_i^* - c_j\right) x_j^* - \mu \sum_{j=1}^{n} x_j^* \ln x_j^* \quad (6.25)$$

另外,按 Q 的定义由(6.21)直接得

$$-\theta_\mu^* = \mu \ln Q$$

以及由 $Ax^*=0$ 推及

$$\sum_{j=1}^{n}\left(\sum_{i=1}^{m} a_{ij} w_i^* - c_j\right) x_j^* = \sum_{i=1}^{m}\left(\sum_{j=1}^{n} a_{ij} x_j^*\right) w_i^* - \sum_{j=1}^{n} c_j x_j^* = -c^T x^*$$

将以上两个关系式一并代入(6.25),得

定理6.15 设 w^* 为(6.21)的解,设 x^* 由(6.24)定义,则 x^* 对(KLP)可行,且对由(6.21)定义的 θ_μ^*,有

$$\theta_\mu^* = c^T x^* + \mu \sum_{j=1}^{n} x_j^* \ln x_j^* \quad (6.26)$$

今再记

$$w_{m+1}^* = \min_{1 \leqslant j \leqslant n}\left\{c_j - \sum_{i=1}^{m} a_{ij} w_i^*\right\} = c_k - \sum_{i=1}^{m} a_{ik} w_i^* \quad (6.27)$$

则易检 $(w_1^*, \cdots, w_m^*, w_{m+1}^*)$ 对(KDP)可行. 根据定理6.11有

$$c^T x^* \geqslant \theta \geqslant w_{m+1}^*$$

又,按(6.27)由(6.24)得 $x_k^* = Q^{-1}\exp(-w_{m+1}^*/\mu)$,在此式两边取

自然对数得

$$\mu \ln x_k^* = -\mu \ln Q - w_{m+1}^* = \theta_\mu^* - w_{m+1}^* \tag{6.28}$$

由于 $x_k^* \in (0,1)$，上式还表明

$$\theta_\mu^* < w_{m+1}^*$$

最后，将(6.26)中 θ_μ^* 的表达式代入(6.28)得

$$c^T x^* - w_{m+1}^* = \mu\left(\ln x_k^* - \sum_{j=1}^{n} x_j^* \ln x_j^*\right) \tag{6.29}$$

定理6.16 设 w^* 为(6.21)的解，或 w^* 由(6.22)定义；θ_μ^* 由(6.21)定义；w_{m+1}^* 由(6.27)定义；x^* 由(6.24)定义，θ 为(KLP)的最优值. 则 x^* 对(KLP)可行，关系式(6.29)成立，并且

$$\theta_\mu^* < w_{m+1}^* \leqslant \theta \leqslant c^T x^* \tag{6.30}$$

由于上式意味着 $0 \leqslant c^T x^* - \theta \leqslant c^T x^* - w_{m+1}^*$，通过对(6.29)右端的详细的分析，我们在[21]中得到了

定理6.17 设 x^* 由(6.24)定义，则 x^* 对(KLP)可行，且

$$0 \leqslant c^T x^* - \theta \leqslant \mu B_n, \ n \geqslant 3, \tag{6.31}$$

其中 θ 为(KLP)的最优值，且

$$B_n = [\ln(n-1) - \ln\ln(n-1)] + \ln[1 + \ln(n-1)] - \ln[\ln(n-1) - \ln\ln(n-1)] - 1, \ n \geqslant 3 \tag{6.32}$$

上定理给出了求解线性规划(KLP)的无约束最优化方法(参见定理6.14下面的一段话). [21]中有例表明，对于这里的方法，定理中的 μB_n 为最好的估计. 在那里还指出了，(6.21)中的极小化问题有解的充要条件为(KLP)有一个可行点 $\hat{x} > 0$.

习　题

1. 设 $C \subseteq R^n$. 证明 C 为凸集的充要条件是，对 C 中的任意有限个点 $x^1, \cdots, x^k$，以及对任意 $\lambda_j \geqslant 0, j = 1, \cdots, k$，其满足 $\sum_{j=1}^{k} \lambda_j = 1$，都有 $\sum_{j=1}^{k} \lambda_j x^j \in C$.

2. 设 S 为 R^n 中的紧集，则 $\mathrm{co}S$ 亦为紧集.

3. 设 C 为 R^n 中的凸集. 证明，$\forall\, x^1 \in \mathrm{cl}C, \forall\, x^2 \in \mathrm{int}C$，

$$y = \lambda x^1 + (1-\lambda)x^2 \in \text{int}C, \forall \lambda \in (0,1)$$

(提示：设 $\epsilon > 0$ 使得 $\{z: \|z - x^2\| < \epsilon\} \subseteq C$，证

$$\{z: \|z - y\| < (1-\lambda)\epsilon\} \subseteq C, \forall \lambda \in (0,1).)$$

4. 设 C 为 R^n 中的凸集. 证明 $\text{int}C$ 和 $\text{cl}C$ 都是凸集.（提示：可利用上题之结果.）

5. 设 C 为 R^n 中的非空凸集，则 $\text{cl}C$ 为所有包含 C 的半闭空间的交.（提示：利用上题之结果及定理6.5.）

6. 设 $A \in R^{m \times n}$，证明

$$D = \{y \in R^m, y = Ax, x \in R^n, x \geqslant 0\}$$

为 R^m 中的非空闭凸集.（提示：关于闭性，先证任意 D 中的 y 都能表成 A 中的若干个线性无关列的非负线性组合.）

7. 设 $A \in R^{m \times n}$, $B \in R^{p \times n}$，证明下面两种情况恰发生一种：

Ⅰ. $Ax < 0, Bx = 0$ 有解 $x \in R^n$；

Ⅱ. $A^T u + B^T v = 0, u \geqslant 0, u \neq 0$ 有解 $(u, v) \in R^{m+p}$.

(提示：必要时可先设 B 的行向量线性无关.)

8. 证明两个凸集之间的严格分离定理：设 D_1 和 D_2 是 R^n 中两个互不相交的非空闭凸集，并设 D_2 是紧的，则存在 $c \in R^n$, $\alpha \in R$，使得

$$\inf_{x \in D_1} c^T x > \alpha > \sup_{y \in D_2} c^T y$$

并指出，若 D_1、D_2 无一为紧则结论一般不成立.

9. 设 $X \subseteq R^n$, $x \in X$. 用 $T(X, x)$ 表 X 在 x 处的切锥：$z \in T(X, x)$ 当且仅当，存在非负数列 $\{\alpha_k\}$ 以及存在 $\{x^k\} \subseteq X$ 且 $x^k \to x$ 使得 $z = \lim \alpha_k (x^k - x)$. 今设 S 为 R^n 中的开凸集且 S 和它的补集 $\mathscr{C}S$ 皆非空. 若 $x \in \text{cl}S \cap \mathscr{C}S$，则

$$(S - \{x\}) \cap T(\mathscr{C}S, x) = \varnothing$$

10. 考虑如下的线性规划：

$$\min \; -(\ln\delta)x_1$$
$$\text{s.t.} \; x_1 - \sum_{j=2}^{n} x_j = 0$$
$$e^T x = 1$$
$$x \geqslant 0$$

其中 $n > 2$, $\delta \in (0,1)$ 为参数.

(i) 验证(6.30)式；

(ii) 试给出 $(c^T x^* - \theta)/\mu$ 的一个对一切 $\delta \in (0,1)$ 通用的上界.

第七章 凸函数

试图将线性规划的理论推广到非线性规划上去，这就导致了对凸函数的详尽的研究．这研究被称为凸分析．凸分析是一门年轻的学科，从20世纪50年代起才开始迅速地发展．本章只限于介绍一些基本理论．

§7.1 凸函数及其基本性质

定义7.1 设 $D\subseteq R^n$ 为凸集，$f:D\to R$．若 $\forall\ x^1,x^2\in D$，

$$f(\alpha x^1+(1-\alpha)x^2)\leqslant \alpha f(x^1)+(1-\alpha)f(x^2) \tag{7.1}$$

对 $\forall\ \alpha\in(0,1)$ 成立，则称 $f(x)$ 为 D 上的凸函数．

若不等式(7.1)对 $x^1\neq x^2$ 严格成立，则称 $f(x)$ 为 D 上的严格凸函数．

又，若 $-f(x)$ 为 D 上的(严格)凸函数，则称 $f(x)$ 为 D 上的(严格)凹函数．

可用归纳法证明，凸函数的一个等价的定义是：对 $\forall\ k,\forall\ x^i\in D,i=1,\cdots,k,\forall\ \alpha_i\geqslant 0,i=1,\cdots,k$ 且 $\sum_{i=1}^k \alpha_i=1$，有

$$f\Big(\sum_{i=1}^k \alpha_i x^i\Big)\leqslant \sum_{i=1}^k \alpha_i f(x^i) \tag{7.2}$$

定理7.2 设 $f_i(x),i=1,\cdots,m$ 为 D 上的凸函数，则 $\forall\ \beta_i\geqslant 0,i=1,\cdots,m$，

$$\sum_{i=1}^m \beta_i f_i(x)$$

亦为 D 上的凸函数．

定理7.3 设 $f_i(x),i\in I$，都是 D 上的凸函数，则

$$f(x)=\sup_{i\in I} f_i(x)$$

亦为 D 上的凸函数.

定理7.4 设 $f_i(x), i\in I$,都是 D 上的凸函数,则 $\forall\ \omega_i, i\in I$,

$$D_\omega = \{x \in D: f_i(x) \leqslant \omega_i, i \in I\}$$

是凸集.

以上的三个定理可用定义直接获证.

定理7.5 设 $f(x)$为开凸集 $D\subseteq R^n$ 上的凸函数,则 $f(x)$在 D 上满足 Lipschitz 条件,即$\forall\ x^0\in D$,存在 x^0的一个邻域 $O(x^0,\varepsilon)$以及常数 L(一般与此邻域有关),使得

$$|f(x^1) - f(x^2)| \leqslant L \cdot \|x^1 - x^2\|, \forall\ x^1, x^2 \in O(x^0,\varepsilon) \tag{7.3}$$

证 先证 $f(x)$是局部有界的.在 D 中选一个以 x^0为中心的立方体 V,顶点为 $v_1, v_2, \cdots, v_m (m=2^n)$.由立方体为顶点的凸包,则 V 中的任一点 x 都可以表为这些顶点的凸组合,即

$$x = \sum_{i=1}^{m}\lambda_i v_i,\ \lambda_i \geqslant 0, \sum_{i=1}^{m}\lambda_i = 1$$

利用凸函数的等价的定义式(7.2),得

$$f(x) \leqslant \sum_{i=1}^{m}\lambda_i f(v_i) \leqslant \max_{1\leqslant i\leqslant m} f(v_i) \triangleq M$$

故 f 在 V 中上有界.另一方面,对 V 中之 x,可选得 V 中之 y,使 $x^0=\frac{1}{2}x+\frac{1}{2}y$.于是

$$f(x^0) \leqslant \frac{1}{2}f(x) + \frac{1}{2}f(y)$$

从而,

$$f(x) \geqslant 2f(x^0) - f(y) \geqslant 2f(x^0) - M$$

故 f 在 V 中下有界.总之,f 在 V 中有界.

再证 f 的局部 Lipschitz 连续性,即(7.3)式.根据上段的结论,对此 x^0,可找到半径为2ε 的球,使 f 在此球形邻域内有界,设界为 K.今任取 $O(x^0,\varepsilon)$中相异的 x^1,x^2,再令

$$x^3 = x^2 + (\varepsilon/\alpha)(x^2 - x^1)$$

其中 $\alpha=\|x^2-x^1\|$.于是 $x^3\in O(x^0,2\varepsilon)$.解出 x^2,得

$$x^2 = \frac{\varepsilon}{\alpha+\varepsilon}x^1 + \frac{\alpha}{\alpha+\varepsilon}x^3$$

再由 f 之凸性得

$$f(x^2) \leqslant \frac{\varepsilon}{\alpha+\varepsilon}f(x^1) + \frac{\alpha}{\alpha+\varepsilon}f(x^3)$$

由此得

$$f(x^2) - f(x^1) \leqslant \frac{\alpha}{\alpha+\varepsilon}[f(x^3) - f(x^1)] \leqslant 2\alpha K/(\alpha+\varepsilon)$$
$$\leqslant 2\alpha K/\varepsilon = (2K/\varepsilon)\|x^2 - x^1\|$$

由于 x^1,x^2可以互换，故又得

$$|f(x^2) - f(x^1)| \leqslant (2K/\varepsilon)\|x^2 - x^1\|$$

取 $L=2K/\varepsilon$，由上式得(7.3)，证毕.

上面这个定理是在20世纪60年代得到的；这里的证明取自[16]，那里还讨论了抽象空间中的有关情况.

下一个定理表明，对多元函数的凸性的判定可化为相关的一维问题. 我们把证明留给读者.

定理7.6 设 $D\subseteq R^n$ 为凸集. f 为 D 上的凸函数的充要条件为，f 限于 D 中的任一条线段上亦为凸函数. 即$\forall\ x^1,x^2\in D$，

$$h(\alpha) = f(\alpha x^1 + (1-\alpha)x^2) \tag{7.4}$$

为[0,1]上的凸函数.

§7.2 凸函数的几个基本定理

定理7.7 设 $C\subseteq R^n$ 为非空凸集，设 $f_i(x)$，$i=1,\cdots,m$，$g_j(x)$，$j=1,\cdots,r$，为 C 上的凸函数，再设 $g_j(x)$，$j=r+1,\cdots,p$ 为线性函数. 若

$$\begin{cases} f_i(x) < 0, i = 1,\cdots,m, g_j(x) \leqslant 0, j = 1,\cdots,r \\ g_j(x) = 0, j = r+1,\cdots,p, x \in C \end{cases} \tag{7.5}$$

无解，则存在 $\alpha_i\geqslant 0$，$i=1,\cdots,m$，$\beta_j\geqslant 0$，$j=1,\cdots,r$，以及 β_j，$j=r+1,\cdots,p$，这些数不全为零，使得

$$\sum_{i=1}^{m}\alpha_i f_i(x) + \sum_{j=1}^{p}\beta_j g_j(x) \geqslant 0, \forall\ x \in C \tag{7.6}$$

证　令

$$Y=\{y\in R^{m+p}:\exists\ x\in C\ 使得\ f_i(x)<y_i,i=1,\cdots,m,$$
$$g_j(x)\leqslant y_{m+j},j=1,\cdots,r,g_j(x)=y_{m+j},j=r+1,\cdots,p\}$$

易知，Y 为非空凸集且不包含零. 利用推论6.8(令那里的 $\bar{x}=0$)知，存在 $\alpha_i,i=1,\cdots,m,\beta_j,j=1,\cdots,p$，这些数不全为零，使得

$$\sum_{i=1}^{m}\alpha_i y_i+\sum_{j=1}^{p}\beta_j y_{m+j}\geqslant 0,\forall\ y\in Y \tag{7.7}$$

由 Y 的构造易得 $\alpha_i\geqslant 0,i=1,\cdots,m,\beta_j\geqslant 0,j=1,\cdots,r$. 比如说，若 $\alpha_1<0$，则在(7.7)中先固定某个 $y\in Y$，再可令 $y_1\to+\infty$，则(7.7)不可能成立，故 $\alpha_1\geqslant 0$. 今任取 $x\in C$，取 $y_i=f_i(x)+\varepsilon,i=1,\cdots,m$，其中 $\varepsilon>0$任意给定；再取 $y_{m+j}=g_j(x),j=1,\cdots,p$，一并代入(7.7)，得

$$\sum_{i=1}^{m}\alpha_i(f_i(x)+\varepsilon)+\sum_{j=1}^{p}\beta_j g_j(x)\geqslant 0,\forall\ x\in C$$

在上式中固定 x 后，令 $\varepsilon\to 0$，得(7.6)，证毕.

在实际应用中，常常要求(7.6)中的诸 α_i 不全为零，于是就有如下的定理，结果取自[17]，证明取自[28].

定理7.8　设 $C\subseteq R^n$ 为非空凸集，设 $f_i(x),i=1,\cdots,m$，为 C 上的凸函数，设 $g_j(x),j=1,\cdots,p$，为线性函数，再设存在 x^0满足

$$g_j(x^0)\leqslant 0,j=1,\cdots,r,g_j(x^0)=0,j=r+1,\cdots,p,\ x^0\in\text{ri}C \tag{7.8}$$

若(7.5)无解，则存在 $\alpha_i\geqslant 0,i=1,\cdots,m$，这些数不全为零，存在 $\beta_j\geqslant 0,j=1,\cdots,r$ 以及 $\beta_j,j=r+1,\cdots,p$，使得(7.6)式成立.

证　对 p 用归纳法. $p=0$时的结果由定理7.7直接推得. 下设 $p\geqslant 1$且设结论对 $p-1$时的情况为真. 若(7.6)中的 $\alpha_i,i=1,\cdots,m$ 不全为零，则结论已得；若在(7.6)中有诸 $\alpha_i=0$，则记

$$J_1=\{j:\beta_j>0,j=1,\cdots,r\}$$
$$J_2=\{j:\beta_j\neq 0,j=r+1,\cdots,p\}$$

由定理7.7知 $J_1\cup J_2\neq\varnothing$，且此时(7.6)化为

$$l(x)\triangleq\sum_{j\in J_1}\beta_j g_j(x)+\sum_{j\in J_2}\beta_j g_j(x)\geqslant 0,\ \forall\ x\in C \tag{7.9}$$

以(7.8)中之 x^0 代入上式得 $l(x^0)=0$. 再由 $l(x)$ 为线性, $l(x^0)=0$, 以及 $x^0\in \mathrm{ri}C$, 由(7.9)容易推得: $l(x)=0, \forall\ x\in C$. 即

$$\sum_{j\in J_1}\beta_j g_j(x)+\sum_{j\in J_2}\beta_j g_j(x)=0, \forall\ x\in C \qquad (7.10)$$

任取定 $j_0\in J_1\cup J_2$, 则在条件(7.10)下, 系统

$$g_j(x)\leqslant 0, j\in J_1, g_j(x)=0, j\in J_2, x\in C$$

的解集合与下系统之解集合相同:

$$g_j(x)=0, j\in J_1\backslash\{j_0\}, g_j(x)=0, j\in J_2\backslash\{j_0\}, x\in C$$

此时, (7.5)无解便等价于

$$\begin{cases} f_i(x)<0, i=1,\cdots,m, g_j(x)\leqslant 0, j\in\{1,\cdots,r\}\backslash J_1 \\ g_j(x)=0, j\in(J_1\cup\{r+1,\cdots,p\})\backslash\{j_0\}, x\in C \end{cases}$$

无解. 注意, 此时问题已化为 $p-1$ 的情况, 根据归纳法假定, 存在 $\hat{\alpha}_i\geqslant 0, i=1,\cdots,m$, 这些数不全为零. 存在 $\hat{\beta}_j\geqslant 0, j\in\{1,\cdots,r\}\backslash J_1$, 以及 $\hat{\beta}_j, j\in(J_1\cup\{r+1,\cdots,p\})\backslash\{j_0\}$, 使得

$$\sum_{i=1}^{m}\hat{\alpha}_i f_i(x)+\sum_{j=1, j\neq j_0}^{p}\hat{\beta}_j g_j(x)\geqslant 0, \forall\ x\in C \qquad (7.11)$$

与所需之结论相比, (7.11)中缺 $g_{j_0}(x)$ 这一项, 并且未必有 $\hat{\beta}_j\geqslant 0, j\in J_1$, 这只要将(7.11)加上(7.10)的适当的正数倍即可. 证毕.

利用定理7.8可以推出许多择一定理. 例如,

定理7.9(Gale)　设 $A\in R^{m\times n}, c\in R^m$, 则下面两种情况恰有一种发生:

Ⅰ. $Ax\leqslant c$　有解 $x\in R^n$;

Ⅱ. $y^TA=0, y^Tc<0, y\geqslant 0$　有解 $y\in R^m$.

证　易证Ⅰ、Ⅱ不能同时成立. 下设Ⅰ不成立. Ⅰ不成立等价于下系统关于 (x,t) 无解:

$$-t<0,\ Ax-ct\leqslant 0$$

由于 $(x,t)=(0,0)$ 满足线性不等式 $Ax-ct\leqslant 0$, 根据定理7.8可知, 存在 $\alpha>0, \alpha\in R, y\geqslant 0, y\in R^m$, 使得

$$-\alpha t+y^T(Ax-ct)\geqslant 0,\forall\, x\in R^n,\forall\, t\in R$$

用 $x=-A^Ty,t=0$ 代入上式得 $y^TA=0$；再在上式中令 $x=0,t=1$ 得 $y^Tc\leqslant-\alpha<0$. 证毕.

定理7.10 设 $C\subseteq R^n$ 为非空凸集，设 $f_i(x),i=1,\cdots,m,$ $g_j(x),j=1,\cdots,s,$ 为 C 上的凸函数，$g_j(x),j=s+1,\cdots,p,$ 为线性函数. 再设对某个 $r:s\leqslant r\leqslant p$，系统

$$\begin{cases}f_i(x)<0,i=1,\cdots,m,g_j(x)\leqslant 0,j=1,\cdots,r\\ g_j(x)=0,j=r+1,\cdots,p,x\in C\end{cases}$$

无解. 若存在 $\hat{x}$ 满足

$$\begin{cases}g_j(\hat{x})<0,j=1,\cdots,s,g_j(\hat{x})\leqslant 0,j=s+1,\cdots,r\\ g_j(\hat{x})=0,j=r+1,\cdots,p,\hat{x}\in \mathrm{ri}C\end{cases}\tag{7.12}$$

则存在 $\alpha_i\geqslant 0,i=1,\cdots,m$，这些数不全为零，存在 $\beta_j\geqslant 0,j=1,\cdots,r$，以及 $\beta_j,j=r+1,\cdots,p$，使得

$$\sum_{i=1}^{m}\alpha_if_i(x)+\sum_{j=1}^{p}\beta_jg_j(x)\geqslant 0,\ \forall\, x\in C\tag{7.13}$$

证 $s=0$ 的情况即定理7.8，下设 $s\geqslant 1$. 由条件利用定理7.8知(7.13)中的 $\alpha_i,i=1,\cdots,m$，及 $\beta_j,j=1,\cdots,s$ 这些数不全为零. 若诸 α_i 皆为零，则 $\beta_j,j=1,\cdots,s$ 不全为零；此时，由(7.12)即得 $\sum_{j=1}^{p}\beta_j\,g_j(\hat{x})<0$，同时由(7.13)又得 $\sum_{j=1}^{p}\beta_j\,g_j(\hat{x})\geqslant 0$，这是一个矛盾. 定理由是得证.

§7.3 凸函数的极值

定理7.11 设 f 为凸集 $D\subseteq R^n$ 上的凸函数，则

(i) f 在 D 上的任一个局部极小点都是全局极小点；

(ii) f 在 D 上的极小点集是凸集；

(iii) 又若 f 为 D 上的严格凸函数，则 f 在 D 上的极小点集或

为空集或为单点集.

证 (i)设 $\bar{x}$ 是 f 在 D 上的一个局部极小点，即存在 $\varepsilon>0$，使得

$$f(x) \geqslant f(\bar{x}), \forall x \in D \cap O(\bar{x},\varepsilon) \tag{7.14}$$

今任取 $y\in D$，则$\forall\ \alpha\in(0,1)$有 $\alpha y+(1-\alpha)\bar{x}\in D$；并且对充分小的 $\alpha>0$，$\alpha y+(1-\alpha)\bar{x}\in O(\bar{x},\varepsilon)$. 于是由(7.14)，对充分小的 $\alpha>0$，就有

$$f(\bar{x}) \leqslant f(\alpha y + (1-\alpha)\bar{x}) \leqslant \alpha f(y) + (1-\alpha)f(\bar{x})$$

容易由上式解得 $f(\bar{x})\leqslant f(y)$. 由 $y\in D$ 的任意性即得

$$f(y) \geqslant f(\bar{x}), \forall y \in D \tag{7.15}$$

此即欲证.

(ii) 设所论的极小点集为 D^*. 任取 $x^1,x^2\in D^*$，则利用(i)可知，$\forall\ \alpha\in(0,1)$，

$$f(\alpha x^1 + (1-\alpha)x^2) \leqslant \alpha f(x^1) + (1-\alpha)f(x^2) = f(x^1) \tag{7.16}$$

并且必有 $f(\alpha x^1+(1-\alpha)x^2)=f(x^1)$，即得 $\alpha x^1+(1-\alpha)x^2\in D^*$，这说明 D^* 为凸集.

(iii) 注意到若 $x^1,x^2\in D^*$ 且 $x^1\neq x^2$，则由 f 之严格凸知(7.16)中之不等式对 $\alpha=\dfrac{1}{2}$严格成立，即 $f\left(\dfrac{1}{2}x^1+\dfrac{1}{2}x^2\right)<f(x^1)$ 这又表明 $x^1\notin D^*$，矛盾. 证毕.

联系着凸函数有一个重要的概念，即次梯度.

定义7.12 设 f 为凸集 $D\subseteq R^n$ 上的凸函数，$\bar{x}\in D$. 若存在 $\xi\in R^n$ 使得

$$f(x) \geqslant f(\bar{x}) + \xi^T(x-\bar{x}), \forall x \in D \tag{7.17}$$

则称 ξ 为 f 在 $\bar{x}$ 处(关于 D)的次梯度.

记 $\partial f(\bar{x})$为由 f 在 $\bar{x}$ 处的次梯度组成的集合. 由定义7.12易得，$\partial f(\bar{x})$为闭凸集，$\forall\ \bar{x}\in D$.

对于给定的 $\bar{x}\in D$，可能 $\partial f(\bar{x})$为空集，可能其为单点集，也可能含有多于一个的点. 例如，令 $D=[-1,1]\subset R$，令

$$f(x)=\begin{cases}|x|, & 若\ x\in[-1,1) \\ 2, & 若\ x=1\end{cases}$$

请读者验证

$$\partial f(x)=\begin{cases}-1, & 若\ x\in[-1,0) \\ [-1,1], & 若\ x=0 \\ 1, & 若\ x\in(0,1) \\ \varnothing, & 若\ x=1\end{cases}$$

能够证明,若 f 为凸集 D 上的凸函数,则 $\forall\ \bar{x}\in \mathrm{int}D$, $\nabla f(\bar{x})$ 存在的充要条件是 $\partial f(\bar{x})$ 为单点集且 $\{\nabla f(\bar{x})\}=\partial f(\bar{x})$. 我们把证明放到习题中去.

若 $f(x)\geqslant f(\bar{x})$, $\forall\ x\in D$, 则令 $\xi=0$ 即知(7.17)成立;反之,若(7.17)对 $\xi=0$ 成立,则有 $f(x)\geqslant f(\bar{x})$, $\forall\ x\in D$. 这个推导过程对 D 和 f 的凸性并没有要求. 但我们有下面的结果.

定理7.13 设 f 是凸集 $D\subseteq R^n$ 上的凸函数,设 S 为非空凸集且 $S\subseteq \mathrm{int}D$. 则 $\bar{x}\in S$ 为 f 在 S 上的极小点的充要条件为,存在 $\xi\in\partial f(\bar{x})$, 即

$$f(x)\geqslant f(\bar{x})+\xi^T(x-\bar{x}),\ \forall\ x\in D$$

且使得

$$\xi^T(x-\bar{x})\geqslant 0,\ \forall\ x\in S$$

证 设 $\xi^T(x-\bar{x})\geqslant 0$, $\forall\ x\in S$ 且 $\xi\in\partial f(\bar{x})$, 则

$$f(x)\geqslant f(\bar{x})+\xi^T(x-\bar{x})\geqslant f(\bar{x}),\ \forall\ x\in S$$

这就表明 $\bar{x}$ 为 f 在 S 上的极小点.

反之,设 $\bar{x}$ 为 f 在 S 上的极小点. 构造 R^{n+1} 中的两个集合如下:

$$C_1=\{(x-\bar{x},y):x\in D,y>f(x)-f(\bar{x})\}$$
$$C_2=\{(x-\bar{x},y):x\in S,y\leqslant 0\}$$

易检,C_1,C_2 都是 R^{n+1} 中的非空凸集,且 $C_1\cap C_2=\varnothing$. 于是利用定理6.9可知,存在 $\xi_0\in R^n$, $u\in R$ 且 $(\xi_0,u)\neq 0$, 存在 $\alpha\in R$, 使得

$$\xi_0^T(x-\bar{x})+uy\geqslant\alpha,\forall\ x\in D,\forall\ y>f(x)-f(\bar{x}) \qquad (7.18)$$

$$\xi_0^T(x-\bar{x})+uy\leqslant\alpha,\forall\ x\in S,\forall\ y\leqslant 0 \qquad (7.19)$$

在(7.19)中令 $x=\bar{x}, y=0$ 得 $\alpha\geqslant 0$. 在(7.18)中，令 $x=\bar{x}$ 得 $uy\geqslant\alpha$，再令 $y\to 0^+$ 得 $\alpha\leqslant 0$，综合得 $\alpha=0$；同时，在 $uy\geqslant\alpha$ 中令 $y\to+\infty$ 又得 $u\geqslant 0$. 若 $u=0$ 则由 $(\xi_0,u)\neq 0$ 推得 $\xi_0\neq 0$. 由于 $\bar{x}\in S\subseteq \mathrm{int}D$，在(7.18)中可取 x 为 $x-\bar{x}=-\lambda\xi_0$，其中 $\lambda>0$ 充分小. 此时，由(7.18)及 $\alpha=0, u=0$ 即得 $\xi_0^T(x-\bar{x})=-\lambda\|\xi_0\|^2\geqslant 0$，由 $\xi_0\neq 0$ 知这是不可能的. 总之，我们证明了 $\alpha=0, u>0$. 用 u 遍除(7.18)，(7.19)，再记 $-\xi=\xi_0/u$，我们得

$$-\xi^T(x-\bar{x})+y\geqslant 0,\ \forall\, x\in D,\ \forall\, y>f(x)-f(\bar{x})$$

$$-\xi^T(x-\bar{x})+y\leqslant 0,\ \forall\, x\in S,\ \forall\, y\leqslant 0$$

在第一式中令 $y\to f(x)-f(\bar{x})$，在第二式中令 $y=0$，分别得欲证之两式. 证毕.

用上面的定理很容易证明

定理7.14 设 f 为开凸集 $D\subseteq R^n$ 上的函数，则 f 为 D 上的凸函数的充要条件为 $\partial f(x)\neq\varnothing, \forall\, x\in D$.

证 若 f 为 D 上的凸函数，则在定理7.13中令 $S=\{x\}, x\in D$ 固定，可知 $\partial f(x)\neq\varnothing$. 反之，任取 $x^1, x^2\in D, \forall\,\alpha\in(0,1)$，且记 $z=\alpha x^1+(1-\alpha)x^2$，任取定 $\xi\in\partial f(z)$，按次梯度之定义式(7.17)，有

$$f(x^1)\geqslant f(z)+\xi^T(x^1-z) \tag{7.20}$$

$$f(x^2)\geqslant f(z)+\xi^T(x^2-z) \tag{7.21}$$

以 α 遍乘(7.20)，以 $(1-\alpha)$ 遍乘(7.21)，将所得之结果相加，得

$$\alpha f(x^1)+(1-\alpha)f(x^2)\geqslant f(z)=f(\alpha x^1+(1-\alpha)x^2)$$

证毕.

§7.4 可微凸函数的性质

定理7.15 设 f 在开凸集 $D\subseteq R^n$ 上连续可微，则 f 是 D 上的凸函数(严格凸函数)的充要条件为，对 $\forall\, x^1, x^2\in D$，有

$$f(x^1)\geqslant f(x^2)+\nabla f(x^2)^T(x^1-x^2) \tag{7.22}$$
$$(>)\qquad\qquad (x^1\neq x^2)$$

证 必要性. 由 f 在 D 上为凸, $\forall\ x^1,x^2\in D,\forall\ \alpha\in(0,1)$,

$$f(\alpha x^1+(1-\alpha)x^2)\leqslant\alpha f(x^1)+(1-\alpha)f(x^2)$$

对上式稍加整理,得

$$f(x^1)-f(x^2)\geqslant\frac{f(x^2+\alpha(x^1-x^2))-f(x^2)}{\alpha}\tag{7.23}$$

在上式中令 $\alpha\to0^+$,得(7.22)之第一式.

若 f 在 D 上为严格凸,令 $z=\frac{1}{2}x^1+\frac{1}{2}x^2$,则

$$f(z)<\frac{1}{2}f(x^1)+\frac{1}{2}f(x^2),x^1\neq x^2$$

又由(7.22)之第一式得

$$f(z)\geqslant f(x^2)+\nabla f(x^2)^T(z-x^2)$$

结合以上两式,得对 $x^1\neq x^2$ 有

$$\begin{aligned}\frac{1}{2}f(x^1)+\frac{1}{2}f(x^2)&>f(x^2)+\nabla f(x^2)^T(z-x^2)\\&=f(x^2)+\frac{1}{2}\nabla f(x^2)^T(x^1-x^2)\end{aligned}$$

由此即可解得(7.22)之第二式.

充分性的证明只要在(7.20)和(7.21)中用 $\nabla f(z)$ 代替那里的 ξ,可以完全类似地证明. 证毕.

定理7.16 设 f 在开凸集 $D\subseteq R^n$ 上二次连续可微. 则 f 为 D 上的凸函数的充要条件为,对 $\forall\ x\in D$, $\nabla^2f(x)$ 为半正定;又若 $\nabla^2f(x)$ 对 $\forall\ x\in D$ 为正定. 则 f 为 D 上的严格凸函数,但反之不然.

证 必要性. 任取 $x\in D$ 及 $d\in R^n$. 因 D 为开集,故对充分小的 α 恒有 $x+\alpha d\in D$. 由定理7.15得

$$f(x+\alpha d)\geqslant f(x)+\alpha\nabla f(x)^Td$$

又因为 f 二次可微,故

$$f(x+\alpha d)=f(x)+\alpha\nabla f(x)^Td+\frac{1}{2}\alpha^2d^T\nabla^2f(x)d+o(\alpha^2)$$

比较上两式得

$$\alpha^2d^T\nabla^2f(x)d+o(\alpha^2)\geqslant0$$

以 α^2 遍除上式后，复令 $\alpha\to 0$ 即得欲证.

充分性. 设 $\nabla^2 f(x)$ 为半正定（正定），$\forall\ x\in D$. 任取 x^1，$x^2\in D$，$x^1\neq x^2$. 由中值定理知，存在 $\hat{x}=x^2+\hat{\alpha}(x^1-x^2)$，$\hat{\alpha}\in(0,1)$，使得

$$
\begin{aligned}
f(x^1) &= f(x^2)+\nabla f(x^2)^T(x^1-x^2) \\
&\quad + \frac{1}{2}(x^1-x^2)^T\nabla^2 f(\hat{x})(x^1-x^2) \\
&\geqslant f(x^2)+\nabla f(x^2)^T(x^1-x^2) \\
&\quad (>)
\end{aligned}
$$

由定理7.15知 f 在 D 上为凸（严格凸）.

最后，考虑 $f(x)=(x)^4$，$x\in R$. 易知 $f(x)$ 在 R 上为严格凸，但 $f''(0)=0$，故由 $f(x)$ 的严格凸推不出 $\nabla^2 f(x)$ 为正定. 证毕.

特别地，若 Q 为 n 阶对称半正定阵，则

$$
f(x)=\frac{1}{2}x^TQx+b^Tx
$$

就是 R^n 上的凸函数. 此时，$\nabla^2 f(x)\equiv Q$，$\forall\ x\in R^n$.

(7.23)右边的商式常常有用. 记

$$
h(\alpha)=f(x+\alpha d)-f(x),\ \alpha>0 \tag{7.24}
$$

若 $f(x)$ 为凸函数，则取 $\alpha_2>\alpha_1>0$，记 $\lambda=\alpha_1/\alpha_2\in(0,1)$，则由定理7.6知

$$
\begin{aligned}
h(\alpha_1) &= h(\lambda\alpha_2)=h(\lambda\alpha_2+(1-\lambda)\cdot 0) \\
&\leqslant \lambda h(\alpha_2)+(1-\lambda)h(0)=\lambda h(\alpha_2)
\end{aligned}
$$

于是就有 $h(\alpha_1)/\alpha_1\leqslant h(\alpha_2)/\alpha_2$，即 $h(\alpha)/\alpha$ 关于 $\alpha>0$ 是单调上升的. 类似地，若由(7.24)定义的 $h(\alpha)$ 中的 f 为凹函数，则 $h(\alpha)/\alpha$ 关于 $\alpha>0$ 是单调下降的. 这个结果在下节会用到.

§7.5 对一类函数的研究

设 $C\subseteq R^m$，$D\subseteq R^n$，$\phi(x,y):C\times D\to R$. 本节考虑函数

$$
g(y)=\inf_{x\in C}\phi(x,y)
$$

的一些性质. 记

$$G(y) = \arg\min_{x\in C}\phi(x,y)$$
$$= \{x \in C: \phi(x,y) = g(y)\}$$

给定 $\bar{y}\in D$, 记满足

$$g(y) - g(\bar{y}) \leqslant \xi^T(y-\bar{y}), \forall\ y \in D$$

的 ξ 的全体为 $\hat{\partial}g(\bar{y})$.

定理7.17 设

1° C 为非空紧集；

2° $\phi(\cdot,y)$ 在 C 上为下半连续；

3° D 为开凸集；

4° $\phi(x,\cdot)$ 为 D 上的凹函数；

5° $\phi'_y(x,y)$ 存在, $\forall\ x\in C$.

则 $g(y)$ 为 D 上的凹函数, 且 $\forall\ y\in D$ 有

$$\hat{\partial}g(y) = \mathrm{cl\ co}\{\phi'_y(x,y): x \in G(y)\} \qquad (7.25)$$

又, 若对某个 $\bar{y}\in D$, $\{\phi'_y(x,\bar{y}): x\in G(\bar{y})\}$ 为单点集, 则 $\nabla g(\bar{y})$ 存在且有表达式

$$\nabla g(\bar{y}) = \phi'_y(x,y)|_{x=x(\bar{y}),y=\bar{y}} \qquad (7.26)$$

其中 $x(\bar{y})\in G(\bar{y})$.

证 利用凹函数的定义由4°可推出 $g(y)$ 为 D 上的凹函数. (7.26)在所论的条件下由(7.25)推出(参见习题8). 故以下只须证(7.25). 注意, 由1°知 $G(y)$ 非空, $\forall\ y\in D$.

任取 $y\in D$, 取 $x^0\in G(y)$. 对 $\forall\ d\in R^n$ 且 $y+d\in D$, 由4°, 5°

$$\begin{aligned} g(y+d) - g(y) &= g(y+d) - \phi(x^0,y) \\ &\leqslant \phi(x^0,y+d) - \phi(x^0,y) \\ &\leqslant \phi'_y(x^0,y)^T d \end{aligned}$$

故得 $\phi'_y(x^0,y)\in \hat{\partial}g(y)$. 由 $\hat{\partial}g(y)$ 为闭凸集又得

$$\mathrm{cl\ co}\{\phi'_y(x,y): x \in G(y)\} \subseteq \hat{\partial}g(y) \qquad (7.27)$$

下证相反的包含关系, 用反证法. 设有 $\xi\in \hat{\partial}g(y)$ 但

$$\xi \notin \mathrm{cl\ co}\{\phi'_y(x,y): x \in G(y)\}$$

由于后者为非空闭凸集, 故由定理6.5知, 存在 $\bar{d}\in R^n$ 使得

$$\xi^T\overline{d} < \eta^T\overline{d}, \forall\ \eta \in \mathrm{cl\ co}\{\phi'_y(x,y): x \in G(y)\} \tag{7.28}$$

另一方面，由 $\xi\in \hat{\partial} g(y)$，我们取正数列 $\{\lambda_i\}$，$\lambda_i \downarrow 0$，再取

$$x^{\lambda_i} \in \arg\min_{x\in C}\phi(x, y + \lambda_i\overline{d}),\ 以及\ x^0 \in G(y)$$

则由4°知，对充分小的 λ_i 有

$$\begin{aligned}
&\lambda_i\phi(x^{\lambda_i}, y + \overline{d}) + (1 - \lambda_i)\phi(x^{\lambda_i}, y) - \phi(x^0, y) \\
&\leqslant \phi(x^{\lambda_i}, y + \lambda_i\overline{d}) - \phi(x^0, y) \\
&= g(y + \lambda_i\overline{d}) - g(y) \\
&\leqslant \lambda_i\xi^T\overline{d}
\end{aligned} \tag{7.29}$$

由1°可不妨设 $x^{\lambda_i}\to x^* \in C$；结合2°，上式表明 $\phi(x^*, y)\leqslant\phi(x^0, y)$；再结合 $x^0\in G(y)$ 得 $x^*\in G(y)$，即 $\phi(x^*, y)=\phi(x^0, y)$.

今任意固定 $\bar{\lambda}\in(0,1)$，则对 $\lambda_i\in(0,\bar{\lambda})$，由(7.29)及4°，

$$\begin{aligned}
\xi^T\overline{d}&\geqslant[\phi(x^{\lambda_i}, y+\lambda_i\overline{d})-\phi(x^0, y)]/\lambda_i \\
&=[\phi(x^{\lambda_i}, y+\lambda_i\overline{d})-\phi(x^{\lambda_i}, y)]/\lambda_i+[\phi(x^{\lambda_i}, y)-\phi(x^*, y)]/\lambda_i \\
&\geqslant[\phi(x^{\lambda_i}, y+\bar{\lambda}\overline{d})-\phi(x^{\lambda_i}, y)]/\bar{\lambda}+[\phi(x^{\lambda_i}, y)-\phi(x^*, y)]/\bar{\lambda}
\end{aligned}$$

上式第二个不等式后的第一项是利用了凹函数的性质(见上节之最后一句)；第二项是利用了关系式 $\phi(x^{\lambda_i}, y)\geqslant\phi(x^*, y)$（因为有了 $x^*\in G(y)$）. 由上式即得

$$\xi^T\overline{d} \geqslant [\phi(x^{\lambda_i}, y + \bar{\lambda}\overline{d}) - \phi(x^*, y)]/\bar{\lambda}$$

在此式中先令 $\lambda_i\to 0$，复令 $\bar{\lambda}\to 0$ 得 $\xi^T\overline{d}\geqslant\phi'_y(x^*, y)^T\overline{d}$，这与(7.28)是矛盾的. 证毕.

定理7.18 设定理7.17中的条件1°～5°全满足. 进一步设 $\phi'(\cdot, y)$是 $G(y)$上的连续函数，则

$$\hat{\partial} g(y) = \mathrm{co}\{\phi'_y(x,y): x \in G(y)\} \tag{7.30}$$

证 由1°、2°知 $G(y)$为非空紧集，再由 $\phi'_y(\cdot, y)$是 $G(y)$上的连续函数知，对固定的 y，$\{\phi'_y(x, y): x\in G(y)\}$也是 R^n 中的非空紧集，而紧集的凸包必为闭集(参见上章习题 2). 于是(7.30)由(7.25)直接得到. 证毕.

文献[3]引进了广义梯度的概念并对上结果作了推广.

引理7.19 设 G 为 R^n 中的凸集，f:$G\to R$. 若对$\forall$ x^1、$x^2\in G$，

$\forall\ \alpha\in(0,1)$，有

$$\alpha f(x^1)+(1-\alpha)f(x^2)=f(\alpha x^1+(1-\alpha)x^2) \quad (7.31)$$

则 f 在 G 上连续.

证 任取 $\{x^k\}\subseteq G, x^k\to x^*\in G$. 若 $f(x^k)\not\to f(x^*)$，则存在 $\varepsilon_0>0$，存在 $\{x^k\}$ 的子集 $\{x^{k_i}\}$ 使得下列两式之一成立：

$$f(x^{k_i})\geqslant f(x^*)+\varepsilon_0, \forall\ k_i \quad (7.32)$$

$$f(x^{k_i})\leqslant f(x^*)-\varepsilon_0, \forall\ k_i \quad (7.33)$$

若(7.32)成立，记 $S=\text{co}\{x^{k_i}\}\subseteq G$，则由(7.31)、(7.32)得

$$f(x)\geqslant f(x^*)+\varepsilon_0, \forall\ x\in S \quad (7.34)$$

今任取一个 $\bar{x}\in\text{ri}S$，因为 $x^*\in\text{cl}S$，故

$$(1-\alpha)x^*+\alpha\bar{x}\in\text{ri}S\subseteq S, \forall\ \alpha\in(0,1)$$

于是，由(7.34)和(7.31)得

$$(1-\alpha)f(x^*)+\alpha f(\bar{x})=f((1-\alpha)x^*+\alpha\bar{x})\geqslant f(x^*)+\varepsilon_0$$

在上式中令 $\alpha\to0$ 得 $\varepsilon_0\leqslant0$，这与 $\varepsilon_0>0$ 矛盾，故(7.32)不成立；同理可证(7.33)也不成立. 这就证得了 f 在 G 上的连续性. 证毕.

定理7.20 设定理7.17中的条件1°～5°全满足，进一步设 C 为凸集且 $\phi(\cdot,y)$ 是 C 上的凸函数，则

$$\hat{\partial}g(y)=\{\phi'_y(x,y):x\in G(y)\} \quad (7.35)$$

证 任取 x^1、$x^2\in G(y)$，$\alpha\in(0,1)$，$d\in R^n$，则对充分小的正数 λ，有

$$\begin{aligned}&\alpha[\phi(x^1,y+\lambda d)-\phi(x^1,y)]+(1-\alpha)[\phi(x^2,y+\lambda d)-\phi(x^2,y)]\\&=[\alpha\phi(x^1,y+\lambda d)+(1-\alpha)\phi(x^2,y+\lambda d)]\\&\quad-[\alpha\phi(x^1,y)+(1-\alpha)\phi(x^2,y)]\\&\geqslant\phi(\alpha x^1+(1-\alpha)x^2,y+\lambda d)-\phi(\alpha x^1+(1-\alpha)x^2,y)\end{aligned}$$

不等式后的第一项是利用了 $\phi(\cdot,y)$ 的凸性假设；其第二项是利用了定理7.11(ii). 以 $\lambda>0$ 遍除上式后，令 $\lambda\to0$，得

$$[\alpha\phi'_y(x^1,y)+(1-\alpha)\phi'_y(x^2,y)]^Td\geqslant\phi'_y(\alpha x^1+(1-\alpha)x^2,y)^Td$$

由于上式对 $\forall\ d\in R^n$ 都成立，故得

$$\begin{aligned}&\alpha\phi'_y(x^1,y)+(1-\alpha)\phi'_y(x^2,y)\\&=\phi'_y(\alpha x^1+(1-\alpha)x^2,y), \alpha\in(0,1) \quad (7.36)\end{aligned}$$

从而就有

$$\begin{aligned}\text{co}\{\phi'_y(x,y):x\in G(y)\} &= \{\phi'_y(x,y):x\in \text{co}G(y)\}\\ &= \{\phi'_y(x,y):x\in G(y)\}\end{aligned}$$

由(7.36)利用引理7.19知 $\phi'(\cdot,y)$在 $G(y)$上为连续；再利用定理7.18知(7.30)成立；最后结合上式知(7.35)成立. 证毕.

习　　题

1. 设 f 为一元凸函数，则对满足

$$a\leqslant b\leqslant c\leqslant d \text{ 且 } d-c=b-a$$

的 a、b、c、d，有

$$f(b)+f(c)\leqslant f(a)+f(d)$$

2. 设 φ 是一个单调非减的单变量函数，它又是凸的，设 f 是定义在凸集 D 上的凸函数. 证明 $\varphi(f(x))$是定义在 D 上的凸函数.

3. 设 g 是 R^n 上的凹函数，则$1/g(x)$为

$$S=\{x:g(x)>0\}$$

上的凸函数；举例说明，若 g 是 R^n 上的凸函数，则$1/g(x)$不必为 S 上的凹函数.

4. 证明定理7.6.

5. 证明非齐次 Farkas 引理：设 $A\in R^{m\times n}$，$b\in R^n$，$c\in R^m$，$q\in R$，则下列两种情况恰发生一种.

Ⅰ. $b^Tx>q$，$Ax\leqslant c$　有解 $x\in R^n$；

Ⅱ. $A^Ty=b$，$c^Ty\leqslant q$，$y\geqslant 0$　有解 $y\in R^m$，或

$A^Ty=0$，$c^Ty<0$，$y\geqslant 0$　有解 $y\in R^m$

6. 设 $c_1,c_2\in R^n$，$\alpha_1,\alpha_2\in R$. 记

$$S=\{x:c_2^Tx+\alpha_2>0\}$$

证明，若 $\bar{x}$ 为

$$f(x)=(c_1^Tx+\alpha_1)/(c_2^Tx+\alpha_2)$$

在 S 上的局部极小，则 $\bar{x}$ 亦为 f 在 S 上的全局极小.

7. 设 f 为凸集 D 上的凸函数且在 D 上不恒为常数. 若 f 在 D 上有最大值，则最大值必在 D 的边界上达到.

8. 设 f 为凸集 D 上的凸函数，则对$\forall\ \bar{x}\in \text{int}D$，$\nabla f(\bar{x})$存在的充要条件为 $\partial f(\bar{x})$为单点集且 $\partial f(\bar{x})=\{\nabla f(\bar{x})\}$.

9. 设 f 为 R^n 上的凸函数，若 f 在 R^n 上有上界，则 f 在 R^n 上必为常数.（提示：可利用次梯度的概念.）

10. 许多有名的不等式能够通过凸性来证明. 试证，对 $a_i \geqslant 0, i = 1,2, \cdots, m$ 且 $\sum_{i=1}^m a_i = 1$，以及对任何的正数 $c_i, i = 1,2,\cdots,m$，有

$$\sum_{i=1}^{m} a_i c_i \geqslant \prod_{i=1}^{m} (c_i)^{a_i}$$

11. 设 f_j 是开集 D 上的二次连续可微的凸函数，则

$$\ln\left[\sum_{j=1}^{m} \exp f_j(x)\right]$$

也是 D 上的凸函数.

12. 设 f_j 是 R^n 上的可微凸函数，记

$$\theta(x) = \max_{1 \leqslant j \leqslant m} \{f_j(x)\}$$

证明 ξ 为 $\theta(x)$ 在 $\bar{x}$ 处的次梯度，当且仅当

$$\xi \in \operatorname{co}\{\nabla f_j(\bar{x}) : f_j(\bar{x}) = \theta(\bar{x}), j = 1,\cdots,m\}$$

13. 设 $a_j > 0, j = 1,\cdots,p$ 且 $\sum_{j=1}^{p} a_j \leqslant 1$. 若 $f_j(x)$ 为凸集 $D \subseteq R^n$ 上取值为正的凹函数，$j = 1,\cdots,p$，则

$$F(x) = \prod_{j=1}^{p} [f_j(x)]^{a_j}$$

也是 D 上的凹函数. 又，若 $\{f_j\}_{j=1}^{p}$ 中有一为 D 上的严格凹函数，则 $F(x)$ 亦为 D 上的严格凹函数.

第八章　可微非线性规划的最优性条件

§8.1　一般形式的最优性条件

考虑如下的一般形式的非线性规划

(P)　　　$\min\limits_{x\in S} f(x)$

其中 $S\subseteq R^n$, $f: R^n\to R$.

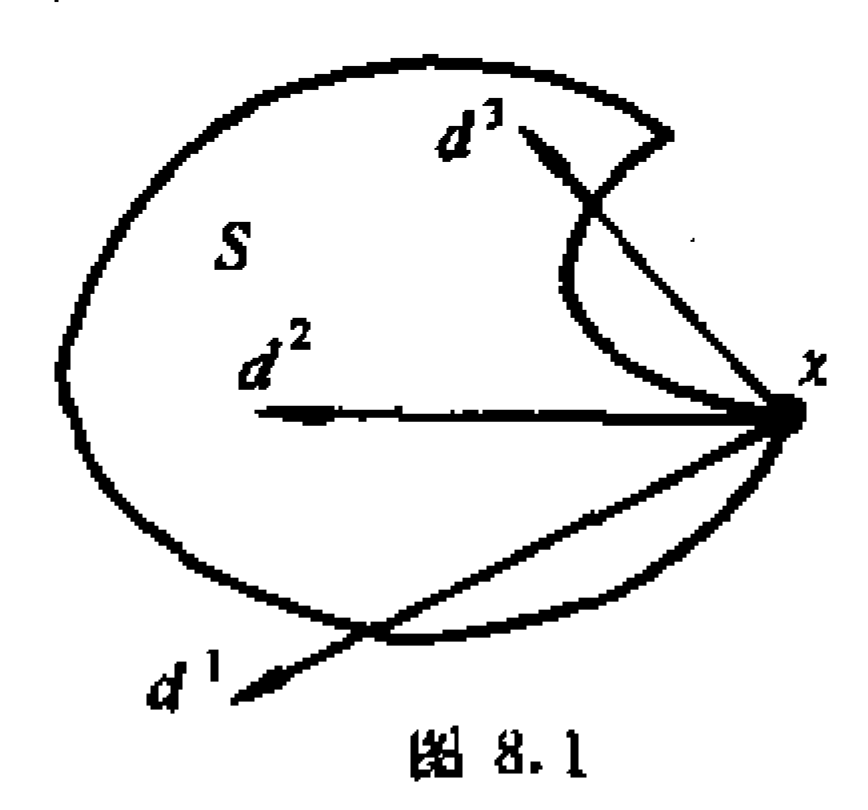


图 8.1

"可行方向"是刻划最优性的重要概念.

定义8.1　设 $x\in S\subseteq R^n$. 说 $d\in R^n$ 是 x 处(关于 S)的一个可行方向,若存在 $\bar{\alpha}>0$,使得

$$x+\alpha d\in S,\ \forall\ \alpha\in[0,\bar{\alpha}]$$

如图8.1,d^1,d^2都是 x 的可行方向,而 d^3则不是.

显见,若 S 为凸集则$\forall\ x,y\in S$,$y-x$ 是 x 处的一个可行方向.

定理8.2　设 f 在包含 $S\subseteq R^n$ 的一个开集上连续可微,若 x^* 是(P)的局部极小点,则对 x^* 处的任一可行方向 d,都有

$$\nabla f(x^*)^Td\geqslant 0$$

证　设 d 为 x^* 处的一个可行方向. 利用 Taylor 公式有

$$f(x^*+\alpha d)=f(x^*)+\alpha\nabla f(x^*)^Td+o(\alpha)$$

若 $\nabla f(x^*)^Td<0$,记 $\beta=-\nabla f(x^*)^Td>0$,则存在 $\varepsilon\in(0,\bar{\alpha})$使得 $|o(\alpha)/\alpha|<\beta/2$,$\forall\ \alpha\in(0,\varepsilon)$. 从而

$$f(x^*+\alpha d)-f(x^*)=\alpha\nabla f(x^*)^Td+o(\alpha)\leqslant-\alpha\beta/2<0 \tag{8.1}$$

对 $\alpha\in(0,\varepsilon)$恒成立. 这说明 x^* 非(P)之局部极小. 证毕.

推论8.3 设 f 在包含 $S\subseteq R^n$ 的一个开集上连续可微. 且设 x^* 为 S 的内点. 若 x^* 为(P)的局部极小点，则 $\nabla f(x^*)=0$.

证 因 x^* 是 S 的内点，故 $\forall\ d\in R^n$ 都是 x^* 处的可行方向. 特别地，在定理8.2中取 $d=-\nabla f(x^*)$，得 $-\|\nabla f(x^*)\|^2\geqslant 0$. 因此必有 $\nabla f(x^*)=0$. 证毕.

推论8.4 设 f 在包含 $S\subseteq R^n$ 的一个开集上连续可微. 进一步设 f 是凸集 S 上的凸函数，则 $x^*\in S$ 是(P)的极小点的充要条件是

$$\nabla f(x^*)^T(x-x^*)\geqslant 0,\ \forall\ x\in S$$

进一步，若 x^* 还是 S 的内点，则 x^* 是(P)的极小点的充要条件是 $\nabla f(x^*)=0$.

证 注意到对 $\forall\ x\in S$，$x-x^*$ 都是 x^* 处的可行方向，必要性由定理8.2立得. 反之，由定理7.15知

$$f(x)\geqslant f(x^*)+\nabla f(x^*)^T(x-x^*),\ \forall\ x\in S$$

从而由条件即得第一个结论. 第二个结论由第一个结论结合推论8.3立得. 证毕.

下面两个定理的证明与定理8.2的证明类似，故留给读者证之.

定理8.5 设 f 在包含 $S\subseteq R^n$ 的一个开集上二次连续可微. 若 x^* 是(P)的局部极小点，则对 x^* 处的任一可行方向 d 有

(i) $\nabla f(x^*)^Td\geqslant 0$;

(ii) 若 $\nabla f(x^*)^Td=0$，则 $d^T\nabla^2 f(x^*)d\geqslant 0$.

定理8.6 设 f 在包含 $S\subseteq R^n$ 的一个开集上二次连续可微，且设 x^* 是 S 的内点. 若 x^* 是(P)的局部极小点，则 $\nabla f(x^*)=0$ 且 $\nabla^2 f(x^*)$ 为半正定.

例8.7 $\min\ f(x_1,x_2)=(x_1)^2-x_1+x_2+x_1x_2$

$\quad$ s.t. $x_1,x_2\geqslant 0$

容易看出，极小点为 $x^*=(1/2,0)^T$. 我们来验证定理8.2和定理8.5. 计算得

$$\nabla f(x^*)=\begin{pmatrix}0\\3/2\end{pmatrix},\quad \nabla^2 f(x^*)=\begin{pmatrix}2&1\\1&0\end{pmatrix}$$

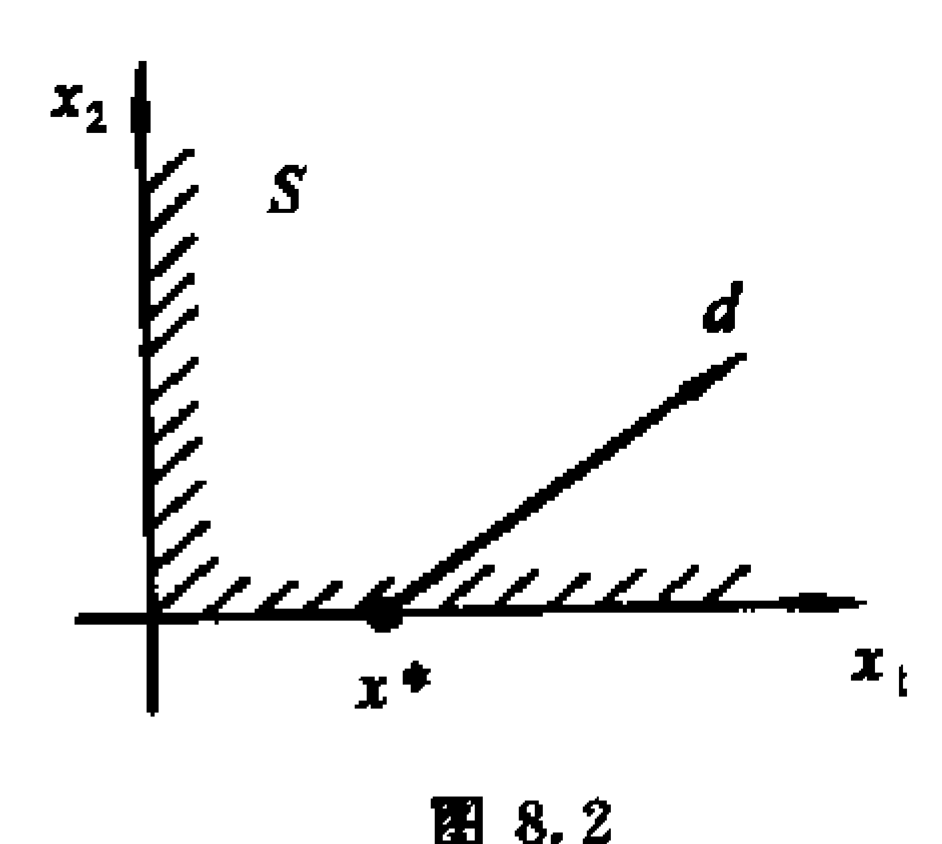


图 8.2

x^* 处的可行方向集为

$$\{d=(d_1,d_2)^T : d_2\geqslant 0\}$$

易检 (i) $\nabla f(x^*)^T d=3/2\cdot d_2\geqslant 0$;

(ii) 对满足 $\nabla f(x^*)^T d=0$ 的 d 形如 $(d_1,0)^T$，于是 $(d_1,0)^T\nabla^2 f(x^*)\begin{pmatrix}d_1\\0\end{pmatrix}=2(d_1)^2\geqslant 0$.

定理8.8 设 f 在包含 $S\subseteq R^n$ 的一个开集上二次连续可微，且设 x^* 是 S 的内点. 若 $\nabla f(x^*)=0$ 且 $\nabla^2 f(x^*)$ 为正定，则 x^* 是(P)的一个严格局部极小点.

证 因 x^* 是 S 的内点，故存在 $\varepsilon_1>0$ 使 $O(x^*,\varepsilon_1)\subseteq S$. 任取 $x\in O(x^*,\varepsilon_1)$，由 $\nabla f(x^*)=0$，利用微分中值定理得

$$f(x)=f(x^*)+\frac{1}{2}(x-x^*)^T\nabla^2 f(x^*+\theta(x-x^*))(x-x^*) \tag{8.2}$$

对某个 $\theta\in(0,1)$ 成立；由 f 的二次连续可微性又知，存在 $\varepsilon_2>0$ 使得 $\forall\ x\in O(x^*,\varepsilon_2)$，$\nabla^2 f(x)$ 为正定. 于是取 $\varepsilon=\min\{\varepsilon_1,\varepsilon_2\}$，从而由(8.2)得

$$f(x)>f(x^*),\ \forall\ x\in O(x^*,\varepsilon)\ 且\ x\neq x^*$$

证毕.

定理8.2和推论8.3说的是一阶必要条件，推论8.4说的是一阶充要条件，定理8.5和定理8.6说的是二阶必要条件，定理8.8说的是二阶充分条件. 由二阶必要条件可以看出，在(P)的极小点 x^* 处目标函数 f 应具有某种凸性(参见定理7.16).

§8.2 标准型的最优性条件

考虑如下的标准型非线性规划：

(NP) $\quad \min f(x)$

$\quad$ s.t. $h_i(x)=0,\ i=1,\cdots,m$

$\quad\quad\quad g_j(x)\leqslant 0,\ j=1,\cdots,p$

其中 $f,h_i,i=1,\cdots,m,g_j,j=1,\cdots,p$ 都是 R^n 上的实值函数. 本节,我们记

$$S=\{x\in R^n:h_i(x)=0,i=1,\cdots,m;g_j(x)\leqslant 0,j=1,\cdots,p\}$$

则(NP)即为 $\min_{x\in S}f(x)$之形式. 故上节的所有结论对此处亦适用. 但对(NP)应发展更有用的最优性条件. 记

$$h=(h_1,\cdots,h_m)^T$$

$$g=(g_1,\cdots,g_p)^T$$

$$X=\{x\in R^n:h(x)=0\}$$

$$\nabla h=(\nabla h_1,\cdots,\nabla h_m)\in R^{n\times m}$$

$$H_x=\{d\in R^n:\nabla h(x)^Td=0\},x\in X$$

以下,为叙述简洁起见,说向量值函数 h(或 g)在 R^n 上为连续可微指的是 h(或 g)的每个分量 h_i(或 g_j)在 R^n 上连续可微,简记为 $h\in C^1$(或 $g\in C^1$),类似地有 $h\in C^2$(或 $g\in C^2$);说 h 是线性函数指的是每个 h_i 为线性;说 g 是凸函数指的是每个 g_j 是凸函数,等等.

引理8.9 设 $x^*\in X$,设 h 在 R^n 上连续可微. 若向量组 $\nabla h_1(x^*),\nabla h_2(x^*),\cdots,\nabla h_m(x^*)$线性无关,则对$\forall\ d\in H_{x^*}$,存在数$\alpha>0$以及一个可微的向量函数 $x(t):[-\alpha,\alpha]\to R^n$,满足

$$x(0)=x^*,h(x(t))=0,|t|\leqslant\alpha \text{ 且 } \dot{x}(0)=d \qquad (8.3)$$

其中 $\dot{x}(t)=(x_1'(t),\cdots,x_n'(t))^T$.

证 任取 $d\in H_{x^*}$. 令 $t\in R,u\in R^m$,考虑以(t,u)为变量的方程组

$$h(x^*+td+\nabla h(x^*)u)=0 \qquad (8.4)$$

由 $x^*\in X$ 知(8.4)有解 $t=0,u=0$. 对(8.4)左边关于 u 求导后,再令$(t,u)=0$,得表达式 $\nabla h(x^*)^T\nabla h(x^*)$,此矩阵非异(因 $\nabla h_1(x^*),\cdots,\nabla h_m(x^*)$线性无关),根据隐函数存在定理知:存在数$\alpha>0$以及一个可微的向量函数 $u(t):|t|\leqslant\alpha,u(0)=0$,并且

$$h(x^*+td+\nabla h(x^*)u(t))=0,|t|\leqslant\alpha \qquad (8.5)$$

对上式关于 t 在 $t=0$处求导,注意到$\nabla h(x^*)^Td=0$,得

$$0=\nabla h(x^*)^T(d+\nabla h(x^*)\dot{u}(0))=\nabla h(x^*)^T\nabla h(x^*)\dot{u}(0)$$

其中 $\dot{u}(t)$的定义如 $\dot{x}(t)$. 仍由$\nabla h(x^*)^T\nabla h(x^*)$非异得 $\dot{u}(0)=0$.

定义

$$x(t) = x^* + td + \nabla h(x^*)u(t), |t| \leqslant \alpha \qquad (8.6)$$

则 $x(0)=x^*$，由(8.5)得 $h(x(t))=0$, $|t|\leqslant\alpha$. 再由(8.6)关于 t 在 $t=0$处求导得

$$\dot{x}(0) = d + \nabla h(x^*)\dot{u}(0) = d$$

故由(8.6)定义之 $x(t)$连续可微且满足(8.3). 证毕.

再记

$$J(x) = \{j: g_j(x) = 0, j = 1, \cdots, p\}$$

注意，对于不等式约束的情况，符号 $J(x)$经常要用到. 以及记

$$G_{x^*} = \{d \in R^n: \nabla g_j(x^*)^T d < 0, j \in J(x^*)\}$$

引理8.10 设 f,h,g 在 R^n 上连续可微，设 $x^* \in S$，即 x^* 对(NP)可行. 若 x^* 是(NP)的局部极小，且 $\nabla h_i(x^*)$, $i=1,\cdots,m$，线性无关，则

$$\{d: \nabla f(x^*)^T d < 0\} \cap H_{x^*} \cap G_{x^*} = \varnothing \qquad (8.7)$$

证 若(8.7)不成立. 取 $\hat{d} \in \{d: \nabla f(x^*)^T d<0\} \cap H_{x^*} \cap G_{x^*}$. 由 $\nabla h_i(x^*)$, $i=1,\cdots,m$，线性无关，利用引理8.9知存在可微函数 $x(t)$满足(8.3). 于是就有

$$\begin{aligned} f(x(t)) &= f(x(0)) + t\nabla f(x(0))^T\dot{x}(0) + o(|t|) \\ &= f(x^*) + t\nabla f(x^*)^T\hat{d} + o(|t|) \qquad (8.8) \end{aligned}$$

类似地，注意到 $g_j(x^*)=0$, $j\in J(x^*)$，又有

$$g_j(x(t)) = t\nabla g_j(x^*)^T\hat{d} + o(|t|), j \in J(x^*) \qquad (8.9)$$

由于 $\hat{d}$ 满足 $\nabla f(x^*)^T\hat{d}<0$, $\nabla g_j(x^*)^T\hat{d}<0$, $j\in J(x^*)$，类(8.1)的推导由(8.8)和(8.9)知，对充分小的 $t>0$，有

$$f(x(t)) < f(x^*),\ g_j(x(t)) < 0,\ j \in J(x^*) \qquad (8.10)$$

又，对 $j\in \{1,\cdots,p\}\backslash J(x^*)$有 $g_j(x^*)<0$，根据 $g_j(x(t))$关于 t 的连续性知，对充分小的 $|t|$，有

$$g_j(x(t)) < 0, j \in \{1,\cdots,p\}\backslash J(x^*) \qquad (8.11)$$

(8.10)与(8.11)一起，再加上 $h(x(t))=0$, $|t|\leqslant\alpha$，表明 x^* 非(NP)的极小点，故得矛盾. 证毕.

定理8.11(Fritz John) 设 f,h,g 在 R^n 上连续可微. 若 x^* 是(NP)的局部极小,则存在 $\lambda_0,\lambda_i,i=1,\cdots,m,\mu_j,j\in J(x^*)$,使得

$$\lambda_0\nabla f(x^*)+\sum_{i=1}^{m}\lambda_i\nabla h_i(x^*)+\sum_{j\in J(x^*)}\mu_j\nabla g_j(x^*)=0 \tag{8.12}$$

$$\lambda_0\geqslant 0,\mu_j\geqslant 0,j\in J(x^*) \tag{8.13}$$

$$(\lambda_0,\lambda,\mu_J)\neq 0 \tag{8.14}$$

其中 $\lambda=(\lambda_1,\cdots,\lambda_m)^T$, μ_J 是由 $\mu_j,j\in J(x^*)$所组成的向量.

注意,(8.12)~(8.14)等价于(8.15)~(8.18):

$$\lambda_0\nabla f(x^*)+\sum_{i=1}^{m}\lambda_i\nabla h_i(x^*)+\sum_{j=1}^{p}\mu_j\nabla g_j(x^*)=0 \tag{8.15}$$

$$\mu_j g_j(x^*)=0,j=1,\cdots,p \tag{8.16}$$

$$\lambda_0\geqslant 0,\mu_j\geqslant 0,j=1,\cdots,p \tag{8.17}$$

$$(\lambda_0,\lambda,\mu)\neq 0 \tag{8.18}$$

其中 $\mu=(\mu_1,\cdots,\mu_p)^T$.

证 若 $\nabla h_i(x^*),i=1,\cdots,m$,线性相关,则存在不全为零的 $\lambda_1,\cdots,\lambda_m$,使 $\sum_{i=1}^{m}\lambda_i\nabla h_i(x^*)=0$. 余下,只要令 $\lambda_0=0,\mu_j=0,j\in J(x^*)$,便知(8.12)~(8.14)成立.

若 $\nabla h_i(x^*),i=1,\cdots,m$,线性无关,则记 A 为由 $\nabla f(x^*)^T$, $\nabla g_j(x^*)^T,j\in J(x^*)$作为行所组成的矩阵,记 B 为由 $\nabla h_i(x^*)^T$, $i=1,\cdots,m$为行所组成的矩阵. 因 x^* 为(NP)为局部极小点,由引理8.10知下系统关于 $d\in R^n$ 无解:

$$Ad<0,\ Bd=0 \tag{8.19}$$

此时,利用第六章习题7的结果直接可验证(8.12)~(8.14)成立. 证毕.

设 x^* 是(NP)的可行点. 若 x^* 能使(8.12)~(8.14)成立,则说 x^* 是(NP)的一个 Fritz John 点. 若对一个具体问题,x^* 是它的 Fritz John 点,但在对应的(8.12)~(8.14)中恰有 $\lambda_0=0$,则这个 Fritz John 点并没有为我们提供 $f(x)$的信息,这不能不说是一个

重大的缺憾。因之，需要寻找条件其能确保在(8.12)～(8.14)中有$\lambda_0>0$，这一类条件被称为约束品性(constraint qualification)。

定理8.12 设f、h，g在R^n上连续可微。若x^*为(NP)的极小点，且下面的Mangasarian-Fromovitz约束品性成立：

$$\begin{cases}\nabla g_j(x^*)^T d<0, j\in J(x^*), \nabla h_i(x^*)^T d=0, i=1,\cdots,m\\ \text{有解 } \hat{d}\in R^n; \nabla h_i(x^*), i=1,\cdots,m, \text{线性无关}\end{cases} \tag{8.20}$$

则存在$\lambda\in R^m$，$\mu\in R^p$，使得

$$\nabla f(x^*)+\sum_{i=1}^m \lambda_i \nabla h_i(x^*)+\sum_{j=1}^p \mu_j \nabla g_j(x^*)=0 \tag{8.21}$$

$$\mu_j g_j(x^*)=0, j=1,\cdots,p \tag{8.22}$$

$$\mu_j\geqslant 0, j=1,\cdots,p \tag{8.23}$$

证 因x^*是(NP)的极小点，故由定理8.11知(8.12)～(8.14)成立。若$\lambda_0=0$，则由(8.12)得

$$\sum_{i=1}^m \lambda_i \nabla h_i(x^*)+\sum_{j\in J(x^*)} \mu_j \nabla g_j(x^*)=0 \tag{8.24}$$

其中$\mu_j\geqslant 0, j\in J(x^*)$，且$\mu_j, j\in J(x^*)$，$\lambda_i, i=1,\cdots,m$，这些数不全为零。以满足(8.20)的$\hat{d}$与上式作内积，得

$$\sum_{i=1}^m \lambda_i \nabla h_i(x^*)^T\hat{d}+\sum_{j\in J(x^*)} \mu_j \nabla g_j(x^*)^T\hat{d}=0$$

若有某个$\mu_j>0$，则上式不能成立，故必有$\mu_j=0, j\in J(x^*)$。于是(8.24)又成为$\sum_{i=1}^m \lambda_i \nabla h_i(x^*)=0$，且$\lambda_i, i=1,\cdots,m$，不全为零，这与$\nabla h_i(x^*), i=1,\cdots,m$，线性无关的假设条件相矛盾。故有$\lambda_0>0$。

以λ_0遍除(8.12)～(8.14)，再将λ_i/λ_0，μ_j/λ_0视为新的λ_i和μ_j。这等价于说：(在$\lambda_0>0$时)在(8.12)～(8.14)中令$\lambda_0=1$，或在(8.15)～(8.18)中令$\lambda_0=1$；而这时的(8.15)～(8.18)又等价于(8.21)～(8.23)。证毕。

若$x^*\in S$能使(8.21)～(8.23)成立，则称x^*为(NP)的Kuhn-Tucker点，简称K-T点，因为(8.21)～(8.23)首由Kuhn和

Tucker 给出（见[10]），对应的(λ,μ)称为与x^*相连带的 K-T 乘子，或广义 Lagrange 乘子. 下一个定理容易证明.

定理8.13 设f,h,g在R^n上连续可微. 若x^*为(NP)的极小点，且

$$\nabla h_i(x^*), i=1,\cdots,m, \nabla g_j(x^*), j\in J(x^*) \text{ 线性无关.} \tag{8.25}$$

则x^*为(NP)的 K-T 点.

寻找在(NP)的极小点x^*处使(8.21)～(8.23)成立的条件，即约束品性，是规划论的一个重要研究方向. 上面的条件(8.20)或(8.25)都是约束品性. 定理8.12和定理8.13的证明思路可概述如下：

x^*是(NP)的极小点	$\Longrightarrow$	x^*是(NP)的 Fritz John 点	$\xrightarrow{\text{某个约束品性}}$	x^*是(NP)的 K-T 点	(8.26)

定理8.14 设f,h,g在R^n上连续可微，且设f,g是R^n上的凸函数，h是线性函数. 若x^*是(NP)的 K-T 点，则x^*是(NP)的全局极小点.

证 由凸性及可微性假设，有

$$f(x)\geqslant f(x^*)+\nabla f(x^*)^T(x-x^*) \tag{8.27}$$

$$h_i(x)=h_i(x^*)+\nabla h_i(x^*)^T(x-x^*), i=1,\cdots,m \tag{8.28}$$

$$g_j(x)\geqslant g_j(x^*)+\nabla g_j(x^*)^T(x-x^*), j=1,\cdots,p \tag{8.29}$$

以λ_i遍乘(8.28)，以μ_j遍乘(8.29)，然后将(8.27)与所得之全部结果相加，得

$$f(x)+\sum_{i=1}^m\lambda_i h_i(x)+\sum_{j=1}^p\mu_j g_j(x)$$

$$\geqslant[f(x^*)+\sum_{i=1}^m\lambda_i h_i(x^*)+\sum_{j=1}^p\mu_j g_j(x^*)]$$

$$+[\nabla f(x^*)+\sum_{i=1}^m\lambda_i\nabla h_i(x^*)+\sum_{j=1}^p\mu_j\nabla g_j(x^*)]^T(x-x^*)$$

上式右端第二项为零（由(8.21)），第一项中$\sum_{j=1}^p\mu_j g_j(x^*)=0$（由(8.22)），$\sum_{i=1}^m\lambda_i h_i(x^*)=0$（因$h_i(x^*)=0, i=1,\cdots,m$）. 于是

上式成为

$$f(x)+\sum_{i=1}^{m}\lambda_i h_i(x)+\sum_{j=1}^{p}\mu_j g_j(x)\geqslant f(x^*),\forall\ x\in R^n$$

若 $x\in S$，则上式左端第二项为零，第三项非正（由(8.23)），故由上式终得

$$f(x)\geqslant f(x^*),\ \forall\ x\in S$$

证毕.

注 若 f,g 是 R^n 上的凸函数，h 是线性函数，则称对应的(NP)为标准凸规划（对 f,g 的可微性这里并不作要求）. 标准凸规划在理论和应用方面都有重要意义. 关于标准凸规划的某些结果可在下章中找到.

例8.15 重写例8.7中的问题

$$\min f(x_1,x_2)=(x_1)^2-x_1+x_2+x_1x_2$$
$$\text{s.t. } x_1,x_2\geqslant 0$$

由于条件(8.25)在此总成立，故问题的极小点必为 K-T 点. 写出对应于(8.21)～(8.23)的式子，以及约束，得

$$\begin{pmatrix}2x_1+x_2-1\\ x_1+1\end{pmatrix}-\mu_1\begin{pmatrix}1\\0\end{pmatrix}-\mu_2\begin{pmatrix}0\\1\end{pmatrix}=0$$

$$\mu_1x_1+\mu_2x_2=0$$

$$\mu_1,\mu_2,x_1,x_2\geqslant 0$$

解之，得唯一解

$$x^*=(1/2,0)^T,\mu_1^*=0,\mu_2^*=3/2$$

由例8.7中的计算知 $\nabla^2 f(x^*)$ 是半正定的，故此例为标准凸规划，从而由定理8.14知此问题有唯一的全局极小点 x^*.

今从与(8.26)不同的思路来寻找约束品性，我们从“可行方向”（见定义8.1）来展开我们的讨论. 仍设 $x^*\in S$. 记

$$D(x^*)=\{d\in R^n: d\text{ 为 }x^*\text{ 处的可行方向}\}\qquad(8.30)$$

若 $d\in D(x^*)$，则存在 $\bar{\alpha}>0$，使得 $\forall\ \alpha\in[0,\bar{\alpha}]$ 成立着

$$h_i(x^*+\alpha d)=0,i=1,\cdots,m$$

$$g_j(x^*+\alpha d)\leqslant 0,j=1,\cdots,p$$

于是由

$$
\begin{aligned}
0 &= h_i(x^* + \alpha d) \\
&= h_i(x^*) + \alpha \nabla h_i(x^*)^T d + o(\alpha) \\
&= \alpha \nabla h_i(x^*)^T d + o(\alpha), i = 1, \cdots, m
\end{aligned}
$$

以 $\alpha>0$ 遍除上式后，复令 $\alpha \to 0$，即得

$$\nabla h_i(x^*)^T d = 0,\ i = 1, \cdots, m \tag{8.31}$$

类似地，由

$$
\begin{aligned}
0 &\geqslant g_j(x^* + \alpha d) \\
&= g_j(x^*) + \alpha \nabla g_j(x^*)^T d + o(\alpha) \\
&= \alpha \nabla g_j(x^*)^T d + o(\alpha), j \in J(x^*)
\end{aligned}
$$

即得

$$\nabla g_j(x^*)^T d \leqslant 0,\ j \in J(x^*) \tag{8.32}$$

再记

$$Z(x^*) = \{d: \nabla h_i(x^*)^T d = 0, i = 1, \cdots, m, \nabla g_j(x^*)^T d \leqslant 0, \\ j \in J(x^*)\} \tag{8.33}$$

则由上面的分析得到

$$D(x^*) \subseteq Z(x^*) \tag{8.34}$$

注意，定理8.2也可表述为：若 x^* 为(NP)的极小点，则

$$\{d: \nabla f(x^*)^T d < 0\} \cap D(x^*) = \varnothing \tag{8.35}$$

设想能将(8.35)中的 $D(x^*)$ 用 $Z(x^*)$ 来代替，则有

$$\{d: \nabla f(x^*)^T d < 0\} \cap Z(x^*) = \varnothing \tag{8.36}$$

顺便指出，条件(8.36)强于条件(8.7). 条件(8.7)等价于 x^* 是(NP)的 Fritz John 点(见定理8.11的证明). 下面的结果表明，条件(8.36)恰等价于 x^* 是(NP)的 K-T 点.

定理8.16 设 f, h, g 在 R^n 上连续可微，$x^* \in S$，则(8.36)成立的充要条件为，存在 $\lambda \in R^m, \mu \in R^p$ 使(8.21)～(8.23)成立. 或用向量的写法就是

$$\nabla f(x^*) + \nabla h(x^*)\lambda + \nabla g(x^*)\mu = 0 \tag{8.37}$$

$$\mu^T g(x^*) = 0 \tag{8.38}$$

$$\mu \geqslant 0 \tag{8.39}$$

其中 ∇g 的定义如 ∇h.

证 记 A 为由 $\nabla h_i(x^*), -\nabla h_i(x^*), i=1,\cdots,m$ 以及 $-\nabla g_j(x^*), j\in J(x^*)$ 作为列所组成的矩阵，记 $b=\nabla f(x^*)$，则(8.36)等价于说

$$凡满足\ d^TA\geqslant 0\ 的\ d\ 必有\ d^Tb\geqslant 0 \tag{8.40}$$

利用 Farkas 引理(定理6.6)容易验证，(8.40)等价于：存在 $\lambda\in R^m$，$\mu_j, j\in J(x^*)$，使得

$$\nabla f(x^*)+\sum_{i=1}^{m}\lambda_i\nabla h_i(x^*)+\sum_{j\in J(x^*)}\mu_j\nabla g_j(x^*)=0 \tag{8.41}$$

$$\mu_j\geqslant 0,\ j\in J(x^*) \tag{8.42}$$

而(8.41)～(8.42)与(8.21)～(8.23)等价. 证毕.

若在(NP)中 h, g 都是线性函数，则对 $\forall\ x^*\in S$ 显然有 $D(x^*)=Z(x^*)$. 这样，由(8.35)即推得(8.36). 从而由定理8.16立得.

定理8.17 设 f 在 R^n 上连续可微，设 h, g 都是线性函数. 若 x^* 是(NP)的极小点，则 x^* 为(NP)的 K-T 点.

Kuhn 和 Tucker 最早建立的约束品性如下：

(KTCQ) 给定 $x^*\in S$. $\forall\ d\in Z(x^*)$，存在定义在$[0,\varepsilon](\varepsilon>0)$上的可微向量函数 $x(t)$，满足

(i) $x(0)=x^*$；

(ii) $x(t)\subseteq S, t\in[0,\varepsilon]$；

(iii) 对某个 $\alpha>0$ 有 $\dot{x}(0)=\alpha d$.

定理8.18(Kuhn-Tucker) 设 f, h, g 在 R^n 上连续可微，设在 $x^*\in S$ 处(KTCQ)成立. 若 x^* 为(NP)的极小点，则 x^* 为(NP)的 K-T 点.

证 因为 x^* 为(NP)的极小点，故 $t=0$必为问题

$$\min_{t\in[0,\varepsilon]} f(x(t))$$

的解. 根据定理8.2(此时1是 $t=0$处的可行方向)得

$$\begin{aligned}0&\leqslant \frac{d}{dt}f(x(t))\big|_{t=0}\cdot 1\\&=\nabla f(x(0))^T\dot{x}(0)\end{aligned}$$

$$= \alpha \nabla f(x^*)^T d, \forall\ d \in Z(x^*)$$

上式表明(8.36)是成立的. 由定理8.16知结论为真. 证毕.

定理8.16和定理8.18的证明思路可概述如下：

$$\boxed{\begin{matrix} x^* \text{是(NP)} \\ \text{的极小点} \end{matrix}} \overset{\text{某个约束品性}}{\Longrightarrow} \boxed{\text{(8.36)成立}} \overset{\text{Farkas引理}}{\Longleftrightarrow} \boxed{\begin{matrix} x^* \text{是(NP)} \\ \text{的 K-T 点} \end{matrix}} \tag{8.43}$$

已经建立了不少的约束品性，这些约束品性和它们之间的关系可参见[2]或[12].

下例表明(NP)的极小点不必为 K-T 点.

例8.19 $\min\ -x_1$

$$\text{s.t.}\ x_2-(1-x_1)^3 \leqslant 0$$

$$x_1, x_2 \geqslant 0, x_1, x_2 \in R$$

易知，$x^*=(1,0)^T$ 是它的全局极小点. 写出对应的(8.37)～(8.39)：

$$\begin{pmatrix} -1 \\ 0 \end{pmatrix} + \mu_1 \begin{pmatrix} 0 \\ 1 \end{pmatrix} - \mu_3 \begin{pmatrix} 0 \\ 1 \end{pmatrix} = 0,\ \mu_1, \mu_3 \geqslant 0$$

这个式子对 μ_1, μ_3是无解的. 请读者验证此例对(8.36)不成立.

利用上述概念可将定理8.14推广如下.

定理8.20 设 f, h, g 在 R^n 上连续可微，设(NP)的可行集 S 为凸集，设 f 是 S 上的凸函数. 若 x^* 是(NP)的 K-T 点，则 x^* 是(NP)的全局极小点.

证 因 x^* 是(NP)的 K-T 点，由定理8.16知(8.36)成立，从而结合(8.34)知(8.35)成立. (8.35)表明对 x^* 处的任一可行方向 d 都有 $\nabla f(x^*)^T d \geqslant 0$. 再利用推论8.4知 x^* 是(NP)的全局极小点. 证毕.

以下讨论二阶最优性条件. 为得到二阶必要条件，我们需要另一个约束品性. 记

$$\begin{aligned} \hat{Z}(x^*) = \{d : & \nabla h_i(x^*)^T d = 0, i = 1, \cdots, m \\ & \nabla g_j(x^*)^T d = 0, j \in J(x^*)\} \end{aligned} \tag{8.44}$$

(MFCQ) 给定 $x^* \in S.\ \forall\ d \in \hat{Z}(x^*)$,存在定义在$[0,\epsilon]$($\epsilon>0$)上的可微向量函数 $x(t)$,满足

(i) $x(0)=x^*$;

(ii) $h_i(x(t))=0, i=1,\cdots,m, g_j(x(t))=0, j\in J(x^*)$,
$t\in[0,\epsilon]$;

(iii) 对某个 $\alpha>0$有 $\dot{x}(0)=\alpha d$.

定理8.21 设 f,h,g 在 R^n 上二次连续可微,设在 $x^*\in S$ 处(MFCQ)成立.若 x^* 是(NP)的极小点且为(NP)的 K-T 点,即存在 $\lambda^*\in R^m, \mu_j^*, j\in J(x^*)$使得(8.21)~(8.23)成立,则对$\forall\ d\in \hat{Z}(x^*)$都有

$$d^T\Big[\nabla^2 f(x^*)+\sum_{i=1}^m \lambda_i^*\nabla^2 h_i(x^*)+\sum_{j\in J(x^*)}\mu_j\nabla^2 g_j(x^*)\Big]d\geqslant 0 \tag{8.45}$$

证 任取 $d\in\hat{Z}(x^*)$,令 $x(t)$是(MFCQ)中的函数.不失一般性,可假定那里的 $\alpha=1$.于是就有

$$\begin{aligned}\frac{\mathrm{d}^2}{\mathrm{d}t^2}h_i(x(0)) &= \dot{x}(0)^T\nabla^2 h_i(x(0))\dot{x}(0)+\ddot{x}(0)^T\nabla h_i(x(0))\\ &= d^T\nabla^2 h_i(x^*)d+\ddot{x}(0)^T\nabla h_i(x^*), i=1,\cdots,m\end{aligned} \tag{8.46}$$

其中 $\ddot{x}(t)=(x_1''(t),\cdots,x_n''(t))^T$.类似地

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}g_j(x(0)) = d^T\nabla^2 g_j(x^*)d+\ddot{x}(0)^T\nabla g_j(x^*), j\in J(x^*) \tag{8.47}$$

从(8.21)~(8.23)和 $\hat{Z}(x^*)$的定义,我们有

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}t}f(x(0)) = d^T\nabla f(x^*)\\ &= -d^T\Big[\sum_{i=1}^m\lambda_i^*\nabla h_i(x^*)+\sum_{j\in J(x^*)}\mu_j^*\nabla g_j(x^*)\Big]=0\end{aligned} \tag{8.48}$$

由于对充分小的 $t\in[0,\epsilon]$,$x(t)$对(NP)可行,以及 x^* 为(NP)的极小点,利用与定理8.18的证明的同样的思路同时利用定理8.5及条件(8.48),我们有 $\mathrm{d}^2 f(x(0))/\mathrm{d}t^2\geqslant 0$,即

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}f(x(0)) = d^T\nabla^2 f(x^*)d + \ddot{x}(0)^T\nabla f(x^*) \geqslant 0 \quad (8.49)$$

用相应的乘子乘(8.46)和(8.47)，然后将所得之结果一起与(8.49)相加，再一次利用(8.21)～(8.23)，即得(8.45). 证毕.

要验证(KTCQ)或(MFCQ)都相当困难，但条件(8.25)就能保证它们都成立. 详见习题11.

为得二阶充分条件，我们要从 $J(x^*)$ 中分出一个子集来. 设 x^* 为(NP)的一个 K-T 点，设 μ^* 是对应于不等式的乘子. 记

$$\hat{J}(x^*) = \{j : j \in J(x^*), \mu_j^* > 0\} \quad (8.50)$$

以及

$$\tilde{Z}(x^*) = \{d : \nabla h_i(x^*)^T d = 0, i = 1,\cdots,m, \nabla g_j(x^*)^T d \leqslant 0,$$
$$j \in J(x^*), \nabla g_j(x^*)^T d = 0, j \in \hat{J}(x^*)\} \quad (8.51)$$

定理8.22 设 f,h,g 在 R^n 上二次连续可微，设 x^* 是(NP)的 K-T 点，即存在 $\lambda^* \in R^m, \mu^* \in R^p$ 使得(8.21)～(8.23)成立. 若对 $\forall\, d \in \tilde{Z}(x^*)$ 且 $d \neq 0$，有

$$d^T\Big[\nabla^2 f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla^2 h_i(x^*) + \sum_{j\in J(x^*)} \mu_j^* \nabla^2 g_j(x^*)\Big]d > 0$$
$$(8.52)$$

则 x^* 为(NP)的一个严格局部极小点.

证 用反证法. 设结论不成立，则存在 S 中的异于 x^* 的序列 $\{x^k\}, x^k \to x^*$，且

$$f(x^k) \leqslant f(x^*),\ k = 1,2,\cdots \quad (8.53)$$

记 $x^k = x^* + \alpha_k d^k$，使得 $\|d^k\| = 1$ 且 $\alpha_k > 0$. 这样就有 $\alpha_k \to 0$. 由 $\|d^k\| = 1$，不妨设 $d^k \to \bar{d}$，并且亦有 $\|\bar{d}\| = 1$.

首先证 $\bar{d} \in \tilde{Z}(x^*)$. 对 $i = 1,\cdots,m$，有

$$0 = h_i(x^k) - h_i(x^*)$$
$$= \nabla h_i(x^* + \xi_{ik}\alpha_k d^k)^T \cdot \alpha_k d^k, 0 \leqslant \xi_{ik} \leqslant 1$$

以 α_k 遍除上式后，令 $k \to \infty$，得

$$\nabla h_i(x^*)^T\bar{d} = 0,\ i = 1,\cdots,m \quad (8.54)$$

类似地有

$$\nabla g_j(x^*)^T\overline{d} \leqslant 0,\ j \in J(x^*) \tag{8.55}$$

以及由(8.53)有

$$\nabla f(x^*)^T\overline{d} \leqslant 0 \tag{8.56}$$

若 $\overline{d} \notin \tilde{Z}(x^*)$，则由(8.54)和(8.55)知，必存在某个 $k \in \hat{J}(x^*)$ 使 $\nabla g_k(x^*)^T\overline{d} < 0$. 但由 x^* 为 K-T 点，我们得到

$$\begin{aligned}\nabla f(x^*)^T\overline{d} &= -\sum_{i=1}^{m}\lambda_i^* \nabla h_i(x^*)^T\overline{d} - \sum_{j\in J(x^*)}\mu_j^* \nabla g_j(x^*)^T\overline{d} \\ &\geqslant -\mu_k^* \nabla g_k(x^*)^T\overline{d} > 0\end{aligned}$$

而这与(8.56)矛盾，故 $\overline{d} \in \tilde{Z}(x^*)$.

其次，利用二阶中值定理得

$$\begin{aligned}0 &\geqslant f(x^k) - f(x^*) \\ &= \nabla f(x^*)^T\alpha_k d^k + \frac{\alpha_k^2}{2}(d^k)^T\nabla^2 f(x^* + \xi_k\alpha_k d^k)d^k \\ &\qquad 0 \leqslant \xi_k \leqslant 1\end{aligned} \tag{8.57}$$

$$\begin{aligned}0 &= h_i(x^k) - h_i(x^*) \\ &= \nabla h_i(x^*)\alpha_k d^k + \frac{\alpha_k^2}{2}(d^k)^T\nabla^2 h_i(x^* + \eta_{ik}\alpha_k d^k)d^k \\ &\qquad 0 \leqslant \eta_{ik} \leqslant 1, i = 1,\cdots,m\end{aligned} \tag{8.58}$$

$$\begin{aligned}0 &\geqslant g_j(x^k) - g_j(x^*) \\ &= \nabla g_j(x^*)\alpha_k d^k + \frac{\alpha_k^2}{2}(d^k)^T\nabla^2 g_j(x^* + \zeta_{jk}\alpha_k d^k)d^k \\ &\qquad 0 \leqslant \zeta_{jk} \leqslant 1, j \in J(x^*)\end{aligned} \tag{8.59}$$

以 λ_i^*，μ_j^* 分别乘以(8.58)和(8.59)，然后将所得之结果一并求和后，再加上(8.57)，再利用 K-T 条件(8.21)～(8.23)，得

$$\begin{aligned}\frac{\alpha_k^2}{2}(d^k)^T\Big\{\nabla^2 f(x^* + \xi_k\alpha_k d^k) + \sum_{i=1}^{m}\lambda_i^* \nabla^2 h_i(x^* + \eta_{ik}\alpha_k d^k) \\ + \sum_{j\in J(x^*)}\mu_j^* \nabla^2 g_j(x^* + \zeta_{jk}\alpha_k d^k)\Big\}d^k \leqslant 0\end{aligned}$$

以 $\alpha_k^2/2$ 遍除上式，然后令 $k\to\infty$，得

$$\bar{d}^T\left[\nabla^2 f(x^*) + \sum_{i=1}^{m}\lambda_i^* \nabla^2 h_i(x^*) + \sum_{j\in J(x^*)} \mu_j^* \nabla^2 g_j(x^*)\right]\bar{d} \leqslant 0$$

因为$\|\bar{d}\|\neq 0$，上式与(8.52)矛盾，定理由是得证.

习　题

1. 证明定理8.5和定理8.6.

2. 求解

$$\begin{aligned} &\max\ -x^2 + 14x - y^2 + 6y \\ &\text{s.t. } x + y \leqslant 2 \\ &\qquad x + 2y \leqslant 3 \end{aligned}$$

3. 给定 $c\in R^n$ 且 $c\neq 0$，考虑

$$\begin{aligned} &\max\ c^T x \\ &\text{s.t. } x^T x \leqslant 1, x \in R^n \end{aligned}$$

证明 $x^* = c/\|c\|$ 是满足二阶充分条件的解.

4. 利用 Kuhn-Tucker 理论求解下面的问题.

$$\begin{aligned} &\min\ (x_1)^2 + x_2 \\ &\text{s.t. } x_1 + x_2 \leqslant 1 \\ &\qquad (x_1)^2 + (x_2)^2 \leqslant 9 \end{aligned}$$

5. 对问题

(IP)
$$\begin{aligned} &\min\ f(x) \\ &\text{s.t. } g_j(x)\leqslant 0,\ j=1,\cdots,p \end{aligned}$$

设 f,g 连续可微，且设 g_j 都是凹函数. 证明若 x^* 是(IP)的极小点，则

$$\{d: \nabla f(x^*)^T d < 0\} \cap \{d: \nabla g_j(x^*)^T d \leqslant 0, j \in J(x^*)\} = \varnothing$$

6. 对可微的标准型非线性规划(NP)，设已有条件(H)使得在此条件下(NP)的任一个 Fritz John 点都是它的全局极小点. 证明，若条件(H)仅与约束有关，则它与"(NP)的可行集至少含有两点"一起就是一个约束品性.

7. 对最小圆问题例5.2，求它的最优性条件并解释其几何意义.

8. 给定正数 $a_i, c_i, i=1,\cdots,n$，以及正数 b. 考虑问题

$$\begin{aligned} &\min\ f(x) = \sum_{i=1}^{n} c_i/x_i \\ &\text{s.t. } \sum_{i=1}^{n} a_i x_i = b \\ &\qquad x_i \geqslant 0, i = 1,\cdots,n \end{aligned}$$

证明目标函数的最优值为

$$f(x^*) = \Big[\sum_{i=1}^{n}(a_ic_i)^{1/2}\Big]^2/b$$

（提示：不必求出 x^*.）

9. 求下面问题的最优值，其中 α 为预先给定的数.

$$\max \sum_{i=1}^{n}\prod_{\substack{j=1\\ j\neq i}}^{n} x_j - \alpha\prod_{j=1}^{n} x_j$$

$$\text{s.t.}\ \sum_{j=1}^{n} x_j = 1$$

$$x_j \geqslant 0,\ j = 1,\cdots,n$$

（提示：可分最优解 x^* 至少有一分量为零及 $x^*>0$两种可能的情况讨论之.）

10. 考虑问题

$$\min f(x_1,x_2) = (x_1-1)^2+(x_2)^2$$

$$\text{s.t.}\ g(x_1,x_2) = x_1-(x_2)^2/\beta \leqslant 0$$

其中 $\beta>0$是参数. 记 $\hat{x}=(0,0)^T$.

(a) 求出 β 的范围，使 K-T 条件在 $\hat{x}$ 处成立；

(b) 求出 β 的范围，使二阶必要条件在 $\hat{x}$ 处成立；

(c) 求出 β 的范围，使二阶充分条件在 $\hat{x}$ 处成立.

11. 在(NP)中，设 $x^*\in S$. 若 g,h 连续可微且

$$\nabla h_i(x^*),i=1,\cdots,m,\nabla g_j(x^*),j\in J(x^*)\ 线性无关 \qquad (8.60)$$

则(KTCQ)和(MFCQ)都成立.（提示：参照引理8.9的证明及引理8.10的证明.）

12. 考虑问题

$$\min_{x\in R^n} f(x) + \omega\|x-a\|$$

其中 $\omega\geqslant 0$，$a\in R^n$ 皆给定. 证明，若 f 为可微凸函数，则 a 为这问题的解的充要条件是

$$\|\nabla f(a)\| \leqslant \omega$$

（提示：可考虑函数$\|x-a\|$在 a 处的次梯度.）

13. 考虑问题

(IP)
$$\min f(x)$$

$$\text{s.t.}\ g_j(x)\leqslant 0,j=1,\cdots,p$$

设 f,g 连续可微. 记

$$S=\{x\in R^n:g_j(x)\leqslant 0,j=1,\cdots,p\}$$

对 $x^* \in S$，记 $J(x^*) = J_1 \cup J_2$，其中

$$\begin{cases} \forall j \in J_1, g_j \text{ 是凹函数} \\ \forall j \in J_2, g_j \text{ 不是凹函数} \end{cases} \tag{8.61}$$

再设

$$\begin{cases} \nabla g_j(x^*)^T d \leqslant 0, j \in J_1 \\ \nabla g_j(x^*)^T d < 0, j \in J_2 \end{cases} \text{有解 } \hat{d} \in R^n \tag{8.62}$$

对 $\forall\, d \in Z(x^*)$，其中

$$Z(x^*) = \{d : \nabla g_j(x^*)^T d \leqslant 0, j \in J(x^*)\}$$

定义

$$x(t) = x^* + t(d + \alpha \hat{d})$$

其中 $\alpha > 0$ 为参数. 则显然

$$x(0) = x^*,\ \dot{x}(0) = d + \alpha \hat{d}$$

证明，(i) 存在 $\varepsilon = \varepsilon(\alpha) > 0$，使得

$$x(t) \subseteq S,\ t \in [0, \varepsilon]$$

(ii) 若 x^* 是(IP)的极小点，则

$$\nabla f(x^*)^T d \geqslant 0,\ \forall\, d \in Z(x^*)$$

(iii) 若 x^* 是(IP)的极小点，则其为(IP)的 K-T 点.

14. 给定 $p \in (1, +\infty)$，$a_j \in R$，$j = 0, 1, \cdots, n$. 求解

$$\begin{aligned} &\min\ \|x\|_p \\ &\text{s.t.}\ \sum_{j=1}^{n} a_j x_j = a_0 \end{aligned}$$

第九章　对偶和鞍点

本章,我们考虑如下的问题:

(P)
$$\min f(x)$$
$$\text{s.t. } g_j(x)\leqslant 0,\ j=1,\cdots,p$$
$$x\in C$$

其中 $C\subseteq R^n$, f,g_j 都是 R^n 上的实值函数.

与上一章的(NP)相比,这里少了等式约束 $h(x)=0$;对本章的大部分推导,只要将 $h(x)=0$化为等价的两个不等式 $h(x)\leqslant 0$ 和 $-h(x)\leqslant 0$,然后将它们归并到不等式约束 $g(x)\leqslant 0$中去,相应的理论即可产生.请读者仔细想一想,这种处理方法对上一章的哪一些理论不适用.又,我们这里还加上了 $x\in C$ 这一约束,这样做会给研究或应用带来方便.例如,若令

$$C=\{x\in R^n:h(x)=0\} \tag{9.1}$$

则(P)就成了前章的(NP).

对于问题(P),定义与之相连的 Lagrange 式如下:

$$\begin{aligned}L(x,\mu)&=f(x)+\sum_{j=1}^{p}\mu_j g_j(x)\\&=f(x)+\mu^T g(x)\end{aligned} \tag{9.2}$$

其中 $x\in C$, $\mu\in R^p$ 且 $\mu\geqslant 0$.

本章的讨论都是与 Lagrange 式有关的.第一、二节所涉及的最优性都是指全局性的.

§9.1　对偶理论

令

$$\theta(\mu)=\inf_{x\in C}L(x,\mu)$$

定义

(D) $$\max \theta(\mu)$$
$$\text{s.t.}\ \ \mu \geqslant 0, \mu \in R^p$$

称(D)为(P)的对偶问题，对应地称(P)为原问题. 记(P)的最优值为$v(P)$，记(D)的最优值为$v(D)$. 确切地说，它们分别是目标函数$f(x)$和$\theta(\mu)$在各自的可行集上的下确界和上确界. 为叙述方便起见，我们假定$v(P)$为有限值.

定理9.1(弱对偶) 对任意对(P)可行的x，任意对(D)可行的μ，有

$$f(x) \geqslant \theta(\mu) \tag{9.3}$$

证 设x, μ分别对(P)和(D)可行，则

$$f(x) \geqslant f(x) + \mu^T g(x) \geqslant \inf_{x \in C} L(x, \mu) = \theta(\mu)$$

此即欲证.

推论9.2 $v(P) \geqslant v(D)$.

证 在(9.3)中先固定x，对$\theta(\mu)$关于μ在R_+^p中求极大，得

$$f(x) \geqslant \max_{\mu \in R_+^p} \theta(\mu) = v(D)$$

既然上式对任意对(P)可行的x都成立，对上式的$f(x)$在可行集上取极小，即得结论.

定理9.3(强对偶) 设$C \subseteq R^n$为凸集，设$f, g_j, j=1,\cdots,r$为C上的凸函数，设$g_j, j=r+1,\cdots,p$为线性函数. 若

$$\begin{cases} g_j(x) < 0, j = 1,\cdots,r, g_j(x) \leqslant 0, \\ j = r+1,\cdots,p \text{ 在 } \mathrm{ri}C \text{ 上有解}, \end{cases} \tag{9.4}$$

则$v(P) = v(D)$.

证 由于$v(P)$是(P)的最优值，故下系统无解

$$f(x) - v(P) < 0, g_j(x) \leqslant 0, j = 1,\cdots,p, x \in C \tag{9.5}$$

由(9.4)利用定理7.10即可推得，存在$\mu_0 > 0, \bar{\mu}_j \geqslant 0, j=1,\cdots,p$，使得

$$\mu_0[f(x) - v(P)] + \sum_{j=1}^{p} \bar{\mu}_j g_j(x) \geqslant 0, \forall\ x \in C \tag{9.6}$$

不失一般性，可在(9.6)中令$\mu_0 = 1$，这样得到

$$f(x)+\sum_{j=1}^{p}\bar{\mu}_j g_j(x)\geqslant v(P),\forall\ x\in C \tag{9.7}$$

从而又推得

$$\theta(\bar{\mu})=\inf_{x\in C}L(x,\bar{\mu})\geqslant v(P)$$

于是更有

$$v(D)=\max_{\mu\in R_+^p}\theta(\mu)\geqslant\theta(\bar{\mu})\geqslant v(P) \tag{9.8}$$

由上式结合推论9.2知 $v(P)=v(D)$. 证毕.

注 从上面的证明可以看出，只要有了(9.7)则必成立 $v(P)=v(D)$. 并且此时就可由(9.8)推出

$$\theta(\bar{\mu})=\max_{\mu\in R_+^p}\theta(\mu) \tag{9.9}$$

这表明 $\bar{\mu}$ 是(D)的解. 于是，我们给出

定义9.4 若存在 $\bar{\mu}_j\geqslant 0, j=1,\cdots,p$，使得

$$f(x)+\sum_{j=1}^{p}\bar{\mu}_j g_j(x)\geqslant v(P),\forall\ x\in C \tag{9.10}$$

则称 $\bar{\mu}=(\bar{\mu}_1,\cdots,\bar{\mu}_p)^T$ 为问题(P)的一个 Kuhn-Tucker 向量.

结合定理9.3及其证明，得

定理9.5 若定理9.3的假设条件成立，则(P)有一个 Kuhn-Tucker 向量；若(P)有一个 Kuhn-Tucker 向量 $\bar{\mu}$，则

$$v(P)=v(D)$$

且 $\bar{\mu}$ 就是(D)的解.

注意，这里的 Kuhn-Tucker 向量和上一章的 Kuhn-Tucker 乘子是两个不同的概念. 下面的定理表明此两概念之间的某种联系.

定理9.6 设 f,g 在 R^n 上连续可微，则

(i) 若(P)有一个 Kuhn-Tucker 向量 $\bar{\mu}$，且(P)有解 $x^*\in\text{int}C$，则

$$\nabla f(x^*)+\sum_{j=1}^{p}\bar{\mu}_j\nabla g_j(x^*)=0 \tag{9.11}$$

$$\bar{\mu}_j g_j(x^*)=0,\ j=1,\cdots,p \tag{9.12}$$

$$\bar{\mu}_j\geqslant 0,\ j=1,\cdots,p \tag{9.13}$$

即 x^* 是(P)的 K-T 点，且 $\bar{\mu}$ 是与 x^* 相连带的 K-T 乘子；反之，

(ii) 若 C 为凸集且 f,g 是 C 上的凸函数，且 x^* 是(P)的 K-T 点，则满足(9.11)～(9.13)的 K-T 乘子 $\bar{\mu}$ 就是(P)的 Kuhn-Tucker 向量.

证 (i)(9.13)显然成立. 用 $f(x^*)$代替(9.10)中的 $v(P)$，得

$$f(x)+\sum_{j=1}^{p}\bar{\mu}_j g_j(x)\geqslant f(x^*),\forall\ x\in C \tag{9.14}$$

在上式中令 $x=x^*$ 得 $\sum_{j=1}^{p}\bar{\mu}_j g_j(x^*)\geqslant 0$；另一方面，由 $\bar{\mu}_j\geqslant 0$，$g_j(x^*)\leqslant 0,j=1,\cdots,p$，又得 $\sum_{j=1}^{p}\bar{\mu}_j g_j(x^*)\leqslant 0$，综合知(9.12)成立. 利用(9.12)可将(9.14)改写为

$$f(x)+\sum_{j=1}^{p}\bar{\mu}_j g_j(x)\geqslant f(x^*)+\sum_{j=1}^{p}\bar{\mu}_j g_j(x^*),\forall\ x\in C \tag{9.15}$$

这就表明 x^* 为函数

$$f(x)+\sum_{j=1}^{p}\bar{\mu}_j g_j(x) \tag{9.16}$$

在 C 上的极小点. 又因为 $x^*\in\mathrm{int}C$，对(9.16)关于 x 在 x^* 求导，其梯度必为零，故得(9.11).

(ii) 由 f,g 为凸且可微，故

$$f(x)\geqslant f(x^*)+\nabla f(x^*)^T(x-x^*),\forall\ x\in C$$

$$g_j(x)\geqslant g_j(x^*)+\nabla g_j(x^*)^T(x-x^*),\forall\ x\in C,j=1,\cdots,p$$

以 $\bar{\mu}_j$ 乘以第二式，$j=1,\cdots,p$，然后将所有结果与第一式一并相加后，再利用(9.11)，即得(9.15). 然后利用(9.12)由(9.15)得(9.14)；注意到(9.14)表明 x^* 为(P)的最小点即 $v(P)=f(x^*)$，以 $v(P)$代替(9.14)中的 $f(x^*)$便完成了我们的证明. 证毕.

下例表明，若(P)没有 Kuhn-Tucker 向量，强对偶定理的结果也可能会成立.

例9.7 设 $x\in R$. 考虑问题

$$\min\ x$$
$$\text{s.t.}\ (x)^2 \leqslant 0$$
$$x \in C = [-1,1]$$

易知 $v(P)=0$. 对应的 Lagrange 式为

$$L(x,\mu) = x + \mu(x^2)$$

易证此例没有 Kuhn-Tucker 向量. 简单的计算表明

$$\theta(\mu) = \min_{x\in C} L(x,\mu) = \begin{cases} -1, & 若\ \mu = 0 \\ -1+\mu, & 若\ 0<\mu\leqslant \dfrac{1}{2} \\ -1/4\mu, & 若\ \mu > \dfrac{1}{2} \end{cases}$$

于是, $\sup_{\mu\geqslant 0}\theta(\mu)=0$. 故 $v(P)=v(D)$.

上例虽不满足定理9.3的条件,但满足下一个定理的条件.

定理9.8(强对偶) 设 $C\subseteq R^n$ 为紧凸集,设 f,g 是 C 上的连续的凸函数. 若(P)的可行解集非空,即存在 $\hat{x}$ 使得

$$g_j(\hat{x}) \leqslant 0,\ j=1,\cdots,p,\hat{x}\in C \tag{9.17}$$

则 $v(P)=v(D)$.

证 任意给定 $\alpha>0$,则下系统无解:

$$f(x) - v(P) + \alpha \leqslant 0, g_j(x) \leqslant 0, j=1,\cdots,p, x\in C \tag{9.18}$$

由 C 为紧集,由上式无解可推及,存在 $\varepsilon>0$使得下系统无解:

$$\begin{cases} f(x) - v(P) + \alpha - \varepsilon \leqslant 0 \\ g_j(x) - \varepsilon \leqslant 0, j=1,\cdots,p, x\in C \end{cases} \tag{9.19}$$

不然,则存在正数列 $\{\varepsilon_k\}$, $\varepsilon_k\to 0$,使得(9.19)在 $\varepsilon=\varepsilon_k$ 时有解 x^k. 因 C 为紧,故 $\{x^k\}$有收敛子列,不妨设 $x^k\to\bar{x}\in C$. 于是在(9.19)中用 ε_k 代替 ε,用 x^k 代替 x 后,令 $k\to\infty$. 由 f,g 的连续性可推出 $\bar{x}$ 为(9.18)之解,得矛盾.

由(9.19)无解利用定理7.7(令那里的 $m=0,p=r$)可知,存在 $\bar{\mu}_j\geqslant 0, j=0,1,\cdots,p$,这些数不全为零,使得

$$\bar{\mu}_0[f(x)-v(P)+\alpha]+\sum_{j=1}^{p}\bar{\mu}_j g_j(x)\geqslant \varepsilon\cdot\sum_{j=0}^{p}\bar{\mu}_j>0$$

$$\forall\, x\in C \tag{9.20}$$

若 $\bar{\mu}_0=0$，则用(9.17)中的 $\hat{x}$ 代入上式便得到矛盾，故 $\bar{\mu}_0>0$.

不妨在(9.20)中令 $\bar{\mu}_0=1$，这样得到

$$f(x)+\sum_{j=1}^{p}\bar{\mu}_j g_j(x)>v(P)-\alpha,\ \forall\, x\in C \tag{9.21}$$

类(9.6)至(9.8)的推导便得

$$v(D)\geqslant v(P)-\alpha \tag{9.22}$$

由于 $\alpha>0$是任给的，故由上式即得 $v(D)\geqslant v(P)$. 由此结合推论9.2即得欲证. 证毕.

注 (9.21)中的 $\bar{\mu}_j$ 一般应与 α 有关，这可从例9.7中得到验证. 从定理9.3和定理9.8中可以看到，强对偶的成立要求有一定的凸性假设. 关于这两个定理在广义分式规划方面的推广，可参见文献[18].

§9.2 鞍点理论

容易看出，

$$\max_{\mu\in R_+^p} f(x)+\sum_{j=1}^{p}\mu_j g_j(x)=\begin{cases}f(x),\text{若 } g_j(x)\leqslant 0, j=1,\cdots,p\\ +\infty,\text{其它情况}\end{cases}$$

于是就有

$$\min_{x\in C}\max_{\mu\in R_+^p} L(x,\mu)=v(P)$$

特别地，若(P)有解 $\bar{x}$，则

$$\max_{\mu\in R_+^p} L(\bar{x},\mu)=\min_{x\in C}\max_{\mu\in R_+^p} L(x,\mu)=v(P) \tag{9.23}$$

又，若(P)有一个 Kuhn-Tucker 向量 $\bar{\mu}$，则由(9.10)得

$$v(D)=\max_{\mu\in R_+^p}\min_{x\in C} L(x,\mu)\geqslant \min_{x\in C} L(x,\bar{\mu})\geqslant v(P)$$

于是又有

$$v(D)=\max_{\mu\in R_+^p}\min_{x\in C}L(x,\mu)=\min_{x\in C}L(x,\bar{\mu}) \qquad (9.24)$$

这是因为有 $v(D)=v(P)$（见定理9.5）. 因之，若(P)有解 $\bar{x}$ 且有一个 Kuhn-Tucker 向量 $\bar{\mu}$，则(9.23)和(9.24)中出现的所有量都互为相等.

能够看出，对$\forall\ \hat{x}\in C$，$\forall\ \hat{\mu}\in R_+^p$，下式平凡成立：

$$\max_{\mu\in R_+^p}L(\hat{x},\mu)\geqslant L(\hat{x},\hat{\mu})\geqslant\min_{x\in C}L(x,\hat{\mu}) \qquad (9.25)$$

从而由上段所述及的条件便可推出

$$\max_{\mu\in R_+^p}L(\bar{x},\mu)=L(\bar{x},\bar{\mu})=\min_{x\in C}L(x,\bar{\mu}) \qquad (9.26)$$

由此终于得到

$$L(\bar{x},\mu)\leqslant L(\bar{x},\bar{\mu})\leqslant L(x,\bar{\mu}),\forall\ x\in C,\forall\ \mu\in R_+^p \quad (9.27)$$

定义9.9 若 $\bar{x}\in C$. $\bar{\mu}\in R_+^p$ 使得(9.27)成立，则称$(\bar{x},\bar{\mu})$为(P)的一个 Lagrange 式的鞍点.

由上面的分析立得

定理9.10 若(P)有解 $\bar{x}$ 且有一个 Kuhn-Tucker 向量 $\bar{\mu}$，则$(\bar{x},\bar{\mu})$就是(P)的一个 Lagrange 式的鞍点.

注 可以仿(9.14)至(9.15)之间的论述证明这个定理，但上面的推导有普遍的意义.

定理9.11 若$(\bar{x},\bar{\mu})$为(P)的 Lagrange 式的鞍点，则 $\bar{x}$ 为(P)的解且 $\bar{\mu}$ 为(P)的一个 Kuhn-Tucker 向量.

证 具体写出(9.27)式为

$$\begin{aligned} f(\bar{x})+\sum_{j=1}^{p}\mu_j g_j(\bar{x}) &\leqslant f(\bar{x})+\sum_{j=1}^{p}\bar{\mu}_j g_j(\bar{x}) \\ &\leqslant f(x)+\sum_{j=1}^{p}\bar{\mu}_j g_j(x) \\ &\forall\ x\in C,\ \forall\ \mu\in R_+^p \end{aligned} \qquad (9.28)$$

由上之第一个不等式得

$$\sum_{j=1}^{p}(\mu_j-\bar{\mu}_j)g_j(\bar{x})\leqslant 0,\ \forall\ \mu\in R_+^p \qquad (9.29)$$

由于对任一个 j_0，取 $\mu_j=\bar{\mu}_j$，$j\neq j_0$，$\mu_{j_0}=\bar{\mu}_{j_0}+1$代入上式得

$g_{j_0}(\bar{x})\leqslant 0$,从而就有

$$g_j(\bar{x}) \leqslant 0,\ j = 1,\cdots,p$$

这表明 $\bar{x}$ 对(P)可行;又,在(9.29)中先令 $\mu_j=0, j=1,\cdots,p$,得 $\sum_{j=1}^{p}\bar{\mu}_j g_j(\bar{x}) \geqslant 0$,复令 $\mu_j = 2\bar{\mu}_j, j = 1,\cdots,p$,得 $\sum_{j=1}^{p}\bar{\mu}_j g_j(\bar{x}) \leqslant 0$,综合得

$$\sum_{j=1}^{p}\bar{\mu}_j g_j(\bar{x}) = 0 \tag{9.30}$$

利用(9.30),(9.28)中的第二个不等式变为

$$f(x) + \sum_{j=1}^{p}\bar{\mu}_j g_j(x) \geqslant f(\bar{x}), \forall\ x \in C \tag{9.31}$$

若 x 对(P)可行,则由(9.31)得 $f(x)\geqslant f(\bar{x})$,这表明 $\bar{x}$ 为(P)之解;此时,(9.31)显然表明 $\bar{\mu}$ 为(P)的 Kuhn-Tucker 向量.证毕.

定理9.10和定理9.11一起表明,$(\bar{x},\bar{\mu})$为(P)的 Lagrange 式鞍点的充要条件为,$\bar{x}$ 为(P)之解且 $\bar{\mu}$ 为(P)的 Kuhn-Tucker 向量.这类结果有的书上也称为不可微规划(即对所涉及的函数的可微性不作要求)的最优性条件.

在本节余下的部分,我们恒设 $f,g\in C^1$.结合定理9.6(ii)和定理9.10可知,若 f,g 是凸集 C 上的凸函数,且 $\bar{x}$ 是(P)的 K-T 点其连带的 K-T 乘子为 $\bar{\mu}$,则这些条件便蕴含着$(\bar{x},\bar{\mu})$是(P)的一个 Lagrange 式鞍点.由此可以看出这种蕴含关系要有一定的凸性假设.现在,我们来给出这种关系成立的充要条件.让我们先复习一下有关概念.注意,以下一般不要求 C 为凸集.

设 C 为开集,考虑

$$\begin{cases} \nabla f(x) + \sum_{j=1}^{p}\mu_j \nabla g_j(x) = 0 \\ \mu_j g_j(x) = 0, j = 1,\cdots,p \\ g_j(x) \leqslant 0, j = 1,\cdots,p \\ \mu_j \geqslant 0, j = 1,\cdots,p \\ x \in C \end{cases} \tag{9.32}$$

设$(\bar{x},\bar{\mu})$为(9.32)的解,则称 $\bar{x}$ 为(P)的 K-T 点.

设 C 为开集，再考虑

$$\begin{cases} \mu_0 \nabla f(x) + \sum_{j=1}^{p} \mu_j \nabla g_j(x) = 0 \\ \mu_j g_j(x) = 0, j = 1, \cdots, p \\ g_j(x) \leqslant 0, j = 1, \cdots, p \\ \mu_0 \geqslant 0, \mu_j \geqslant 0, j = 1, \cdots, p \\ (\mu_0, \mu) \neq 0 \\ x \in C \end{cases} \tag{9.33}$$

若$(\bar{x}, \bar{\mu}_0, \bar{\mu})$为(9.33)的解，则称$\bar{x}$为(P)的 Fritz John 点.

再考虑

$$f(\bar{x}) + \sum_{j=1}^{p} \mu_j g_j(\bar{x}) \leqslant f(\bar{x}) + \sum_{j=1}^{p} \bar{\mu}_j g_j(\bar{x}) \leqslant f(x) + \sum_{j=1}^{p} \bar{\mu}_j g_j(x), \quad \forall x \in C, \forall \mu \in R_+^p \tag{9.34}$$

设(9.34)的解为$(\bar{x}, \bar{\mu}) \in C \times R_+^p$. 则称$(\bar{x}, \bar{\mu})$为(P)的 Lagrange 式的鞍点.

最后考虑

$$\bar{\mu}_0 f(\bar{x}) + \sum_{j=1}^{p} \mu_j g_j(\bar{x}) \leqslant \bar{\mu}_0 f(\bar{x}) + \sum_{j=1}^{p} \bar{\mu}_j g_j(\bar{x}) \leqslant \bar{\mu}_0 f(x) + \sum_{j=1}^{p} \bar{\mu}_j g_j(x), \quad \forall x \in C, \forall \mu \in R_+^p \tag{9.35}$$

设(9.35)的解为$(\bar{x}, \bar{\mu}_0, \bar{\mu}) \in C \times R_+ \times R_+^p$ 且$(\bar{\mu}_0, \bar{\mu}) \neq 0$，则称$(\bar{x}, \bar{\mu}_0, \bar{\mu})$为(P)的 Fritz John 式鞍点.

记：

W_{KTP}：(9.32)的解$(\bar{x}, \bar{\mu})$的全体；

W_{KTSP}：(9.34)的解$(\bar{x}, \bar{\mu})$的全体；

W_{FJP}：(9.33)的解$(\bar{x}, \bar{\mu}_0, \bar{\mu})$的全体；

W_{FJSP}：(9.35)的解$(\bar{x}, \bar{\mu}_0, \bar{\mu})$的全体.

注 K-T 点和 Fritz John 点的定义要求 C 为开集；但这里的 W_{KTP}和 W_{FJP}的定义并不要求 C 为开集.

接下来，我们提出条件(K)：

(K) 存在 $\eta: C\times X\to R^n$，使得只要 $x\in C$ 及 $\bar{x}\in X$，就有

$$\begin{cases} f(x)-f(\bar{x})-\nabla f(\bar{x})^T\eta(x,\bar{x})\geqslant 0 \\ g_j(x)-\nabla g_j(\bar{x})^T\eta(x,\bar{x})\geqslant 0, j\in J(\bar{x}) \end{cases} \tag{9.36}$$

其中

$$X=\{x\in C: g_j(x)\leqslant 0, j=1,\cdots,p\}$$

$$J=J(\bar{x})=\{j: g_j(\bar{x})=0, j=1,\cdots,p\}$$

注意，若 f,g 都是凸集 C 上的凸函数，则在(9.36)中置 $\eta=x-\bar{x}$，即知(K)成立.

定理9.12 (K)成立的充要条件为

$$W_{FJP}\subseteq W_{FJSP} \tag{9.37}$$

证 必要性的证明留给读者. 下证充分性. 任意取定 $x\in C$ 及 $\bar{x}\in X$，用 $g_J(x)$表由$\{g_j(x)\}_{j\in J(\bar{x})}$组成的列向量；用$\nabla g_J(\bar{x})$表由$\{\nabla g_j(\bar{x})\}_{j\in J(\bar{x})}$作为列所组成的矩阵. 将下面的式子

$$A=\begin{bmatrix}\nabla f(\bar{x})^T \\ \nabla g_J(\bar{x})^T\end{bmatrix}, c=\begin{pmatrix} f(x)-f(\bar{x}) \\ g_J(x)\end{pmatrix}, x:=\eta(x,\bar{x}) \tag{9.38}$$

代入以下的两个式子中：

$$Ax\leqslant c \text{ 有解 } x \tag{9.39}$$

$$A^Ty=0, c^Ty<0, y\geqslant 0 \text{ 有解 } y \tag{9.40}$$

Gale 定理(定理7.9)说，(9.39)和(9.40)恰有一成立. 容易看出，(9.39)成立便是条件(K)成立. 于是以下只须证在条件(9.37)下(9.40)不成立. 用反证法.

设 $y=(\xi_0,\xi_J)$为(9.40)的解，其中 ξ_J 为由$\{\xi_j\}_{j\in J(\bar{x})}$组成的向量. 由(9.40)中的 $y\geqslant 0$及 $c^Ty<0$得$(\xi_0,\xi_J)\neq 0$且$(\xi_0,\xi_J)\geqslant 0$；再结合(9.40)中的 $A^Ty=0$知$(\bar{x},\xi_0,\xi)\in W_{FJP}$(其中 $\xi=(\xi_1,\cdots,\xi_p)^T$：在 $j\in J$ 时取 ξ_j 为 ξ_J 中的对应的元，在 $j\notin J$ 时取 $\xi_j=0$). 从而由条件(9.37)成立推出$(\bar{x},\xi_0,\xi)\in W_{FJSP}$，从而(9.35)的第二个不等式

对$(\bar{\mu}_0,\bar{\mu})=(\xi_0,\xi)$成立. 但这又与(9.40)中的$c^Ty<0$矛盾. 从而(9.40)不成立. 证毕.

定理9.13 若(P)的每一个Fritz John点都是它的K-T点，则(K)成立的充要条件为

$$W_{\mathrm{KTP}}\subseteq W_{\mathrm{KTSP}} \tag{9.41}$$

注 若C为开集，则上两定理中的包含关系都可以改为等号. 例如，可参见定理9.6(i)和定理9.11的证明.

关于定理9.13的推广和应用可参见文献[19]，与之有关的一些结果可参见文献[7,14].

§9.3 Lagrange式的局部凸化

本节，我们恒设$f,g\in C^2$且C为开集. 设$\bar{x}$是(P)的K-T点，其相连带的K-T乘子为$\bar{\mu}$，即$(\bar{x},\bar{\mu})$是(9.32)的解. 由于$g_j(\bar{x})\leqslant 0,j=1,\cdots,p$，故$\sum_{j=1}^{p}\mu_jg_j(\bar{x})\leqslant 0,\forall\ \mu\in R_+^p$. 所以，

$$f(\bar{x})+\sum_{j=1}^{p}\mu_jg_j(\bar{x})\leqslant f(\bar{x})+\sum_{j=1}^{p}\bar{\mu}_jg_j(\bar{x}),\forall\ \mu\in R_+^p$$

这样，(9.34)的第一个不等式平凡成立. 为使$(\bar{x},\bar{\mu})$为(P)的Lagrange式鞍点，问题的焦点成为寻找条件使得

$$f(x)+\sum_{j=1}^{p}\bar{\mu}_jg_j(x)\geqslant f(\bar{x})+\sum_{j=1}^{p}\bar{\mu}_jg_j(\bar{x}),\forall\ \mu\in C \tag{9.42}$$

或

$$L(x,\bar{\mu})\geqslant L(\bar{x},\bar{\mu}),\forall\ x\in C \tag{9.43}$$

注意到(9.32)的第一式为$\nabla_xL(\bar{x},\bar{\mu})=0$，使(9.43)成立的一个充分条件为$L(x,\bar{\mu})$关于$x$在$C$上是凸的(补设$C$的凸集)，这个条件等价于$\nabla_{xx}^2L(x,\bar{\mu})$在$C$上为半正定. 若(P)不是凸规划，直接考察$\nabla_{xx}^2L(x,\bar{\mu})$的半正定性会有所困难. 退一步考虑，若$\nabla_{xx}^2L(\bar{x},\bar{\mu})$为正定，则依$f,g$的二次连续可微性即推得

$\nabla^2_{xx}L(x,\bar{\mu})$在 $\bar{x}$ 的某个邻域内为正定，从而结合$\nabla_x L(\bar{x},\bar{\mu})=0$即知(9.43)局部成立，即存在$\epsilon>0$，使得

$$L(x,\bar{\mu}) \geqslant L(\bar{x},\bar{\mu}), \forall\ x \in C \cap O(\bar{x},\epsilon) \qquad (9.44)$$

通过对一般的规划问题(P)进行等价变换，得一个新问题的 Lagrange 式 $\hat{L}(x,\mu)$，同时 $\bar{x}$ 仍是新问题的 K-T 点，设与之相连带的 K-T 乘子为 $\hat{\mu}$. 我们能够证明，若(P)在 $\bar{x}$ 处的某种二阶最优性充分条件成立，则$\nabla^2_{xx}\hat{L}(\bar{x},\hat{\mu})$为正定. 将 $\hat{L}(x,\mu)$视为(P)的新的 Lagrange 式，这样便完成了对(P)的 Lagrange 式的局部凸化.

今考虑非线性规划

(P) $\qquad \min f(x)$

$$\text{s.t. } g_j(x)\leqslant 0, j=1,\cdots,p$$
$$x\in C$$

对 $\bar{x}\in C,\bar{\mu}\in R^p_+$，记

$$J(\bar{x}) = \{j: g_j(\bar{x}) = 0, j = 1,\cdots,p\}$$

$$\hat{J}(\bar{x}) = \{j: \bar{\mu}_j > 0, j \in J(\bar{x})\}$$

$$\nabla_x L(\bar{x},\bar{\mu}) = \nabla f(\bar{x}) + \sum_{j=1}^{p}\bar{\mu}_j\nabla g_j(\bar{x}) \qquad (9.45)$$

$$\nabla^2_{xx}L(\bar{x},\bar{\mu}) = \nabla^2 f(\bar{x}) + \sum_{j=1}^{p}\bar{\mu}_j\nabla^2 g_j(\bar{x}) \qquad (9.46)$$

$$\bar{Z}(\bar{x}) = \{d \in R^n: \nabla g_j(\bar{x})^T d = 0, j \in \hat{J}(\bar{x})\}$$

容易看出，(P)等价于

(P'_a) $\qquad \min f(x)$

$$\text{s.t. } \exp[\alpha g_j(x)]\leqslant 1, j=1,\cdots,p$$
$$x\in C$$

其中 $\alpha>0$为参数. (P'_a)的 Lagrange 式为

$$\hat{L}(x,\mu) = f(x) + \sum_{j=1}^{p}\mu_j\{\exp[\alpha g_j(x)] - 1\} \qquad (9.47)$$

今计算得

$$\nabla_x\hat{L}(\bar{x},\hat{\mu})=\nabla f(\bar{x})+\sum_{j=1}^{p}\hat{\mu}_j\cdot\alpha\nabla g_j(\bar{x})\cdot\exp[\alpha g_j(\bar{x})] \qquad (9.48)$$

$$\nabla^2_{xx}\hat{L}(\bar{x},\hat{\mu})=\nabla^2 f(\bar{x})+\sum_{j=1}^{p}\hat{\mu}_j\cdot\alpha\nabla^2 g_j(\bar{x})\cdot\exp[\alpha g_j(\bar{x})]$$

$$+\sum_{j=1}^{p}\hat{\mu}_j\cdot\exp[\alpha g_j(\bar{x})]\cdot\alpha^2\nabla g_j(\bar{x})\nabla g_j(\bar{x})^T \quad (9.49)$$

今定义 $\hat{\mu}=(\hat{\mu}_1,\cdots,\hat{\mu}_p)^T$(它与 α 有关)为

$$\hat{\mu}_j=\begin{cases}\bar{\mu}_j/\alpha, 若\ j\in J(\bar{x})\\ 0, 若\ j\notin J(\bar{x})\end{cases} \quad (9.50)$$

由(9.48)和(9.49)结合(9.46),注意到

$$\exp[\alpha g_j(\bar{x})]=1, j\in J(\bar{x})$$

得

$$\nabla_x\hat{L}(\bar{x},\hat{\mu})=\nabla f(\bar{x})+\sum_{j=1}^{p}\bar{\mu}_j\nabla g_j(\bar{x}) \quad (9.51)$$

$$\begin{aligned}\nabla_{xx}^2\hat{L}(\bar{x},\hat{\mu})&=\nabla^2 f(\bar{x})+\sum_{j=1}^{p}\bar{\mu}_j\nabla^2 g_j(\bar{x})\\&\quad+\alpha\cdot\sum_{j=1}^{p}\bar{\mu}_j\nabla g_j(\bar{x})\nabla g_j(\bar{x})^T\\&=\nabla_{xx}^2 L(\bar{x},\bar{\mu})+\alpha\cdot\sum_{j=1}^{p}\bar{\mu}_j\nabla g_j(\bar{x})\nabla g_j(\bar{x})^T \quad (9.52)\end{aligned}$$

定理9.14 对于问题(P),设 $(\bar{x},\bar{\mu})$ 是(9.32)的解,且由(9.46)定义的 $\nabla_{xx}^2 L(\bar{x},\bar{\mu})$ 在 $\bar{Z}(\bar{x})$ 上为正定,则对由(9.47)定义的 $\hat{L}(x,\mu)$ 以及由(9.50)定义的 $\hat{\mu}$ 有 $\nabla_x\hat{L}(\bar{x},\hat{\mu})=0$;并且存在 $\alpha_0>0$ 使得对 $\forall\ \alpha\geqslant\alpha_0$, $\nabla_{xx}^2\hat{L}(\bar{x},\hat{\mu})$ 为正定.

证 由 $(\bar{x},\bar{\mu})$ 满足(9.32),对照(9.51)和(9.45)即得

$$\nabla_x\hat{L}(\bar{x},\hat{\mu})=0$$

下证第二个结论,用反证法.设结论不成立,即存在单调上升的正数列 $\{\alpha_k\}$, $\alpha_k\to+\infty$,存在 $\{d^k\}\subseteq R^n$, $d^k\neq 0$,使得

$$(d^k)^T\nabla_{xx}^2\hat{L}(\bar{x},\hat{\mu})d^k\leqslant 0$$

此即,利用(9.52),

$$(d^k)^T\nabla_{xx}^2 L(\bar{x},\bar{\mu})d^k+\alpha_k\cdot\sum_{j=1}^{p}\bar{\mu}_j\|\nabla g_j(\bar{x})^T d^k\|^2\leqslant 0 \quad (9.53)$$

在上式中,不妨设 $\|d^k\|=1$ 及 $d^k\to\bar{d}$,于是 $\bar{d}\neq 0$. 再在上式中令 $\alpha_k\to+\infty$,从而推得

$$\nabla g_j(\bar{x})^T\bar{d}=0, j\in\hat{J}(\bar{x})$$

这样就有

$$\bar{d} \in \bar{Z}(\bar{x}) \text{ 且 } \bar{d} \neq 0 \tag{9.54}$$

另一方面，由(9.53)直接推出$(d^k)^T \nabla^2_{xx} L(\bar{x},\bar{\mu}) d^k \leqslant 0$，从而

$$\bar{d}^T \nabla^2_{xx} L(\bar{x},\bar{\mu}) \bar{d} \leqslant 0 \tag{9.55}$$

(9.55)和(9.54)一起表明$\nabla^2_{xx} L(\bar{x},\bar{\mu})$在$\bar{Z}(\bar{x})$上不是正定的，这与原设矛盾. 证毕.

关于局部鞍点和局部凸化的进一步的结果可参见文献[22].

习 题

1. 考虑问题

$$\text{(P)} \qquad \begin{aligned} \min\ & (x_1)^2+(x_2)^2 \\ \text{s.t. }\ & x_1+x_2 \geqslant 4 \\ & x_1, x_2 \geqslant 0 \end{aligned}$$

求解(P)和(D).

2. 对于如下的线性规划

$$\text{(LP)} \qquad \begin{aligned} \min\ & c^T x \\ \text{s.t. }\ & Ax \geqslant b \\ & x \geqslant 0 \end{aligned}$$

建立它的对偶问题和对偶理论(要求利用本章的结果).

3. 考虑问题

$$\text{(P)} \qquad \begin{aligned} \min\ & f(x) \\ \text{s.t. }\ & g_j(x) \leqslant 0, j \in J \\ & x \in C \end{aligned}$$

其中J为无限集. 这种规划称为半无限规划. 设C为紧凸集，$f, g_j, j \in J$都是C上的连续凸函数. 试将定理9.8推广到这里的问题. (提示：对每一个$j \in J$，定义集合

$$\Gamma_j = \{x \in C : f(x) - v(P) + \alpha \leqslant 0, g_j(x) \leqslant 0\}$$

其中$\alpha > 0$为给定，$v(P)$为(P)的最优值. 这表明$\bigcap_{j \in J} \Gamma_j = \varnothing$；由于每个$\Gamma_j$都是紧集，可利用有限交性质找到$J$的一个有限不空的子集$J_0$使得$\bigcap_{j \in J_0} \Gamma_j = \varnothing$.)

4. 构造一个规划(P)，使得$v(P)$和$v(D)$都有限. 且有$v(P) > v(D)$. 此时我们说(P)和(D)之间有一个对偶间隙.

5. 设 $\phi(x,y): C\times D\to R$，则对 $\bar{x}\in C$ 和 $\bar{y}\in D$

$$\phi(\bar{x},y)\leqslant\phi(\bar{x},\bar{y})\leqslant\phi(x,\bar{y}),\forall\, x\in C.\forall\, y\in D$$

成立的充要条件是

$$\max_{y\in D}\min_{x\in C}\phi(x,y)=\phi(\bar{x},\bar{y})=\min_{x\in C}\max_{y\in D}\phi(x,y)$$

6. 对于问题

(IP) $\min f(x)$

$\text{s.t. } g_j(x)\leqslant 0, j=1,\cdots,p$

考虑它的 Wolfe 式对偶：

(WD) $\max\limits_{y,\mu} f(y)+\sum\limits_{j=1}^{p}\mu_j g_j(y)$

$\text{s.t. } \nabla f(y)+\sum\limits_{j=1}^{p}\mu_j\nabla g_j(y)=0$

$\mu_j\geqslant 0, j=1,\cdots,p$

(i) 试给出(IP)和(WD)之间弱对偶成立的充要条件；

(ii) 设两规划之间的弱对偶成立. 若(IP)有解且在某个解 x^* 处某个约束品性成立，则(WD)亦有解且两规划的最优值相同.

7. 证明定理9.12中的必要性.

8. 证明定理9.13.（提示：应对"若(P)的每一个 Fritz John 点都是它的 K-T 点"这句话作精细的理解.）

9. 对于问题

$$\begin{aligned}\min\ & -(x+1)/(1-x)\\ \text{s.t. } & x\leqslant 0\\ & x\in C=\{x\in R: x<1\}\end{aligned}$$

(i) 证明它对于 Lagrange 式 $L(x,\mu)$ 没有鞍点；

(ii) 对于 Lagrange 式 $\hat{L}(x,\mu)$，找一个 $a_0>0$ 使得 $\forall\, a\geqslant a_0$，它都有鞍点并指出这鞍点；

(iii) 对(ii)中的任一个鞍点 $(\bar{x},\hat{\mu})$，证明 $\bar{x}$ 是原问题的解.

10. 对于分式规划

(FP) $\min f(x)/d(x)$

$\text{s.t. } g_j(x)\leqslant 0, j=1,\cdots,p$

$x\in C$

其中 $d(x)>0,\forall\, x\in C$. 设其最优值 $v(FP)$ 为有限. 对(FP)定义它的 Lagrange 式如下

$$L(x,\mu)=f(x)/d(x)+\sum_{j=1}^{p}\mu_j g_j(x)/d(x),x\in C,\mu\in R_+^p$$

利用 $L(x,\mu)$建立分式规划的对偶理论和鞍点理论.

(提示:可从下系统无解

$$f(x)-v(FP)d(x)<0,g_j(x)\leqslant 0,j=1,\cdots,p,x\in C$$

出发,在一定的凸性设定下进行推导.)

11. 研究 R^3中如下的问题:

$$\min \cos x_1+\cos x_2+\cos x_3$$
$$\text{s.t. } x_1+x_2+x_3=\pi$$
$$x_1,x_2,x_3\geqslant 0$$

验证在它的任一极小点处,定理9.14中的结论不成立.此例表明定理9.14中的条件"$\nabla_{xx}^2L(\bar{x},\bar{\mu})$在 $\bar{Z}(\bar{x})$上为正定"中的"正定"不能减弱为"半正定".

12. 利用§7.5的结果对§9.1所定义的函数 $\theta(\mu)$(其中 $\mu\in R^m$)的次梯度进行研究.再证明,若在 $\theta(\mu)$的定义式中取 $C=R^n$,取 f 为带有正定阵 Q 的二次函数,取 g 为线性函数,则 $\theta(\mu)$是可微的,同时写出 $\nabla\theta(\mu)$的表达式.

13. 设 $B\in R^{m\times n}$且$B\neq 0$及 $\eta\in R^m$ 为给定,考虑极大化问题

(P):

$$\max \eta^T Bx$$
$$\text{s.t. } \|Bx\|_p=1,\ x\in R^n$$

以及极小化问题

(Q):

$$\min \|\xi\|_q$$
$$\text{s.t. } B^T\xi=B^T\eta,\ \xi\in R^m$$

其中 $p,q>1$且$\frac{1}{p}+\frac{1}{q}=1$亦为给定.试证这两个问题都有最优解且它们的最优值相同;对后一结论作几何解释.(注:这个结论对 $p=1$和 $q=+\infty$,或对 $p=+\infty$和 $q=1$的情况亦成立.)

第十章　基本的下降法

应该指出，利用前两章所述的最优性条件一般只能求解低维空间中的较为简单的问题．因此，构造求解非线性规划问题的算法就成为重要的了．这个内容通常称为最优化方法．

构造算法的一个简单想法是：对$\min\limits_{x\in S} f(x)$，先构造一个映射$A:S\to S$，再寻找一个初始点$x^0\in S$，利用$A$作迭代$x^{k+1}=A(x^k)$，$k=0,1,2,\cdots$，使得$\{f(x^k)\}$为单调下降序列．故称这种算法为下降算法．也即

算法10.0

步0　找一个$x^0\in S$，$k:=0$；

步1　$x^{k+1}=A(x^k)$；

步2　若$f(x^{k+1})\geq f(x^k)$，停；否则$k:=k+1$，返步1

于是，立即会出现下面三个问题，它们构成最优化的主要内容．

(i) 如何构造映射A；

(ii) 如果算法10.0在第k步停于步2，x^k能提供什么信息；又若算法产生一个无穷点列$\{x^k\}$，它的极限点（如果有的话）能提供什么信息．这问题称为收敛性；

(iii) 如果$\{x^k\}$收敛到某点x^*，收敛的速度如何，用什么尺度来衡量．

第十、十一章讨论无约束最优化，第十二、十三章讨论约束最优化．

§10.1　全局收敛性

为使讨论的问题更具一般性，我们先定义概念，然后将算法10.0加以推广．从而讨论推广了的算法具备何种条件可确保收敛

性.本节的结果适用于约束和无约束最优化.

定义10.1 设 $\Gamma\subseteq R^n$ 是一个解集合,设 A 是一个点到集的映射:$X\to Y(X,Y\subseteq R^n)$,设 ω 是一个已知的 R^n(或 $X\cup Y$)中的实值连续函数,说 ω 关于 Γ 和 A 是下降的,若

(i) $\forall\ x\notin\Gamma,\forall\ y\in A(x)$,有 $\omega(y)<\omega(x)$;

(ii) $\forall\ x\in\Gamma,\forall\ y\in A(x)$,有 $\omega(y)\leqslant\omega(x)$.

例10.2 $\min x^2$, s. t. $x\in R$

定义 $\Gamma=\{0\}$,$A(x)=[-|x|/2,|x|/2]$,取 $\omega=(x)^2$. 容易检证 ω 关于 Γ 和 A 是下降的.

算法10.3(下降算法)

设 Γ,A 如定义10.1,其中 $X\supseteq Y$.

步0 任取 $x^0\in X$,$k:=0$;

步1 任取 $x^{k+1}\in A(x^k)$;

步2 若 $x^{k+1}\in\Gamma$,停;否则 $k:=k+1$,返步1.

将算法10.3用于例10.2以及那里的 Γ,A. 容易看出,或有某个 $x^k=0$,或算法给出无穷序列 $\{x^k\}$,且 $x^k\to0$. 这就是算法的收敛性. 下面定义的收敛性其涵义更广.

定义10.4 设算法10.3产生了一个无穷序列 $\{x^k\}$. 若 $\{x^k\}$ 有一个聚点属于 Γ,即存在 $\{x^k\}$ 的一个收敛子列 $\{x^{k_i}\}$ 满足 $x^{k_i}\to x^*\in\Gamma$,则说算法10.3是全局收敛的.

研究算法10.3的收敛性对于构造各种具体的算法及其收敛性的证明具有普遍的指导意义. 算法的收敛与否与映射 A 的性质密切相关. 于是就有

定义10.5 设 $X\subseteq R^n$,$Y\subseteq R^m$. 说点到集的映射 $A:X\to Y$ 在 $x\in X$ 处是闭的,如果由

(i) $x^k\in X,x^k\to x$;

(ii) $y^k\in A(x^k),y^k\to y$,

可以推出 $y\in A(x)$. 又,若 A 在 X 上的任一点处都是闭的,则称 A 在 X 上是闭的.

特别地,若 $A:R^n\to R$ 是连续实值函数,则由定义10.5,A 显然

是闭的. 故闭映射可以看成是连续函数的推广.

有了以上的准备，我们可以给出

定义10.6（全局收敛性定理） 设解集合 Γ 为预先给定，设 $A:X\to Y(Y\subseteq X\subseteq R^n)$ 是点到集的映射，设算法10.3产生一个无穷点列 $\{x^k\}$. 若

(i) 存在一个 X 上的连续实值函数 ω，它关于 Γ 和 A 是下降的；

(ii) 存在一个紧集 $S\subseteq X$，使得 $\{x^k\}\subseteq S$；

(iii) 映射 A 在 $X\backslash\Gamma$ 上是闭的，

则算法10.3是全局收敛的.

证 设 $\{x^{k_i}\}$ 是 $\{x^k\}$ 的一个子列，$x^{k_i}\to x^*$. 由算法知 $\{\omega(x^k)\}$ 是单调下降的，且由 ω 的连续性知 $\omega(x^{k_i})\to\omega(x^*)$，$i\to\infty$. 从而亦有 $\omega(x^k)\to\omega(x^*)$，$k\to\infty$.

下证 $x^*\in\Gamma$. 若 $x^*\notin\Gamma$，先由 $\{x^{k_i+1}\}\subseteq S$ 及 S 为紧知 $\{x^{k_i+1}\}$ 有一个收敛的子列 $\{x^{(k_i)_j+1}\}$，设其收敛于 $\bar{x}$，同时显然有 $x^{(k_i)_j}\to x^*$. 由 $x^{(k_i)_j+1}\in A(x^{(k_i)_j})$ 及 A 在 x^* 处为闭的条件(iii)，得 $\bar{x}\in A(x^*)$. 于是由 ω 关于 A 和 Γ 为下降得到 $\omega(\bar{x})<\omega(x^*)$. 另一方面，由上一段的陈述知必有 $\omega(\bar{x})=\omega(x^*)$. 这个矛盾便说明了 $x^*\in\Gamma$. 证毕.

再考察例10.2中的 Γ,A,ω，易知定理10.6中的条件全满足，故算法10.3用之必收敛. 由定理10.6的证明可以看出映射 A 为闭的重要性，因此有必要对它们作进一步研究.

定义10.7 设 $A:X\to Y$，$B:Y\to Z$ 都是点到集的映射.（其中 X,Y,Z 可以视为不同空间中的子集）合成映射 $C=BA:X\to Z$ 定义为 $C(x)=B(A(x))=\bigcup\limits_{y\in A(x)}B(y)$.

合成映射可用图10.1示之，它是复合函数的概念的推广. 下一定理给出复合映射为闭的一些充分条件.

定理10.8 设 $A:X\to Y$，$B:Y\to Z$ 是点到集的映射. 假定 A 在 x 处是闭的，再假定 B 在 $A(x)$ 上是闭的且若 $x^k\to x$，$y^k\in A(x^k)$，则 $\{y^k\}$ 有一收敛的子序列，那么合成映射 $C=BA$ 在 x 处是闭的.

证 设 $x^k \to x, x^k \in X, z^k \in C(x^k), z^k \to z$，我们来证 $z \in C(x)$.

由 $z^k \in C(x^k) = BA(x^k)$，可选出 $y^k \in A(x^k)$ 使得 $z^k \in B(y^k)$. 据假定，$\{y^k\}$ 有子列 $y^{k_i} \to y$. 由 A 在 x 处为闭知 $y \in A(x)$，再由 B 在 $A(x)$ 上闭（自然在点 y 亦闭）推得 $z \in B(y) \subseteq B(A(x)) = C(x)$. 证毕.

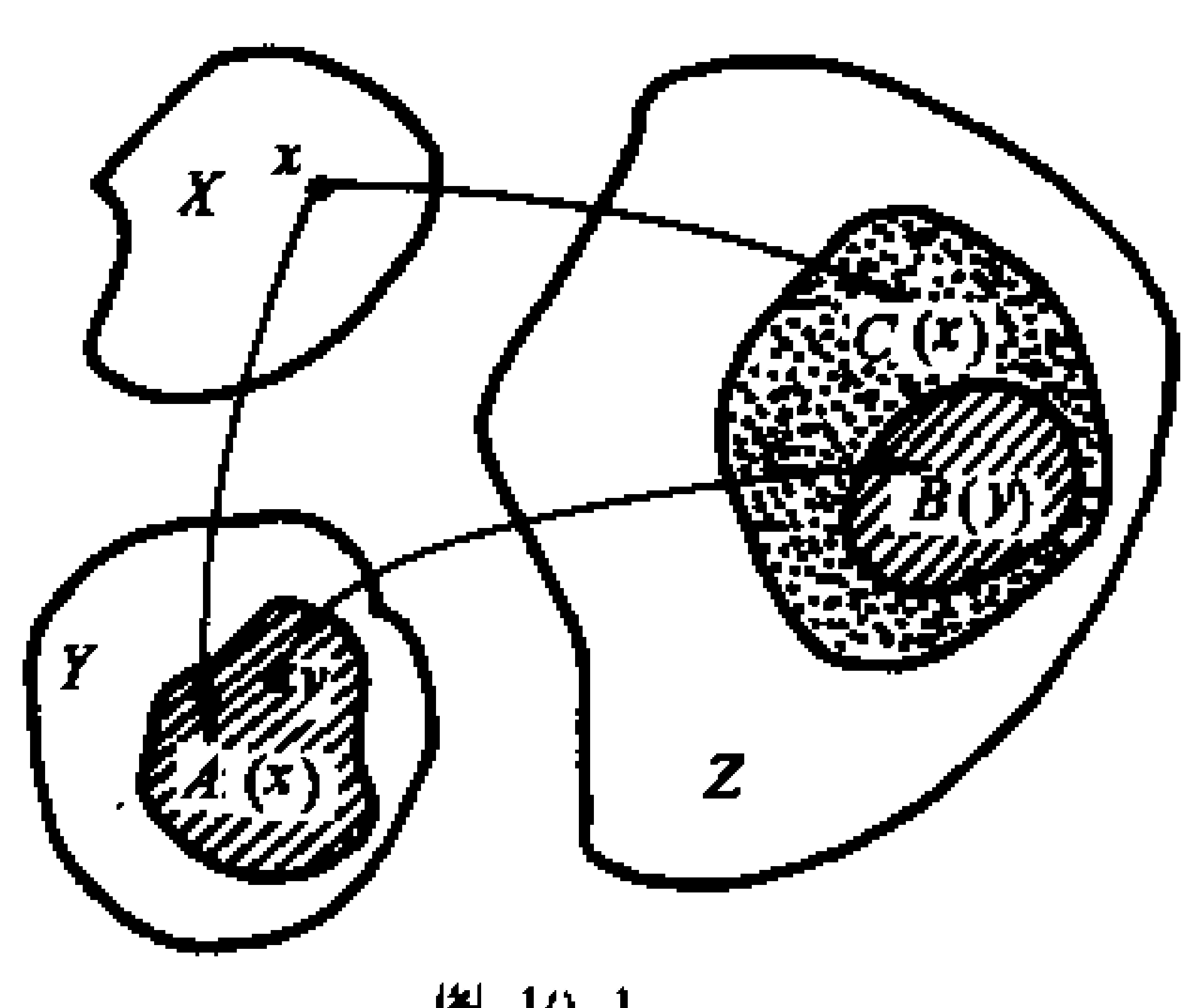


图 10.1

推论10.9 设 $A: X \to Y, B: Y \to Z$ 是点到集的映射. 如果 A 在 x 处是闭的，B 在 $A(x)$ 上闭的且 Y 是紧的，则 $C = BA$ 在 x 处是闭的.

推论10.10 设 $A: X \to Y$ 是点到点的映射，$B: Y \to Z$ 是点到集的映射. 如果 A 在 x 处连续且 B 在 $A(x)$ 处为闭，则 $C = BA$ 在 x 处是闭的.

最后简单提一下收敛速度的概念.

定义10.11 设 $\{x^k\}$ 为收敛点列，$x^k \to x^*$.

(i) 若存在 q 及 N，使得只要 $k > N$，就有

$$\|x^{k+1} - x^*\| \leqslant q\|x^k - x^*\|^p \tag{10.1}$$

则称 $\{x^k\}$ 为 p 次收敛，或称收敛的阶为 p.

(ii) 特别地，若在(10.1)中有 $p=1$ 且 $q<1$，则称 $\{x^k\}$ 为线性收敛；在(10.1)中若 $p=2$，则称 $\{x^k\}$ 为二次收敛.

(iii) 若对 $\forall\, q>0$，存在 $N=N(q)$，使得只要 $k>N$ 就有

$$\|x^{k+1} - x^*\| \leqslant q\|x^k - x^*\| \tag{10.2}$$

则称 $\{x^k\}$ 为超线性收敛.

超线性收敛的一个等价的定义为

$$\limsup_{k\to\infty} \|x^{k+1} - x^*\| / \|x^k - x^*\| = 0 \tag{10.3}$$

由定义即知：

p 次收敛($p>1$)$\Rightarrow$超线性收敛$\Rightarrow$线性收敛.

§ 10.2 一维最优化

本节研究如何寻找定义在某直线或某线段或某射线上的函数的极小点. 此称为一维最优化,或称线搜索.

1. Fibonacci 法和黄金分割法

Fibonacci 法和黄金分割法适用于单峰函数. 单峰函数定义如下.

定义10.12 设 $\varphi(\alpha)$是定义在$[a,b]$上的函数. 称 $\varphi(\alpha)$为$[a,b]$上的下单峰函数,若

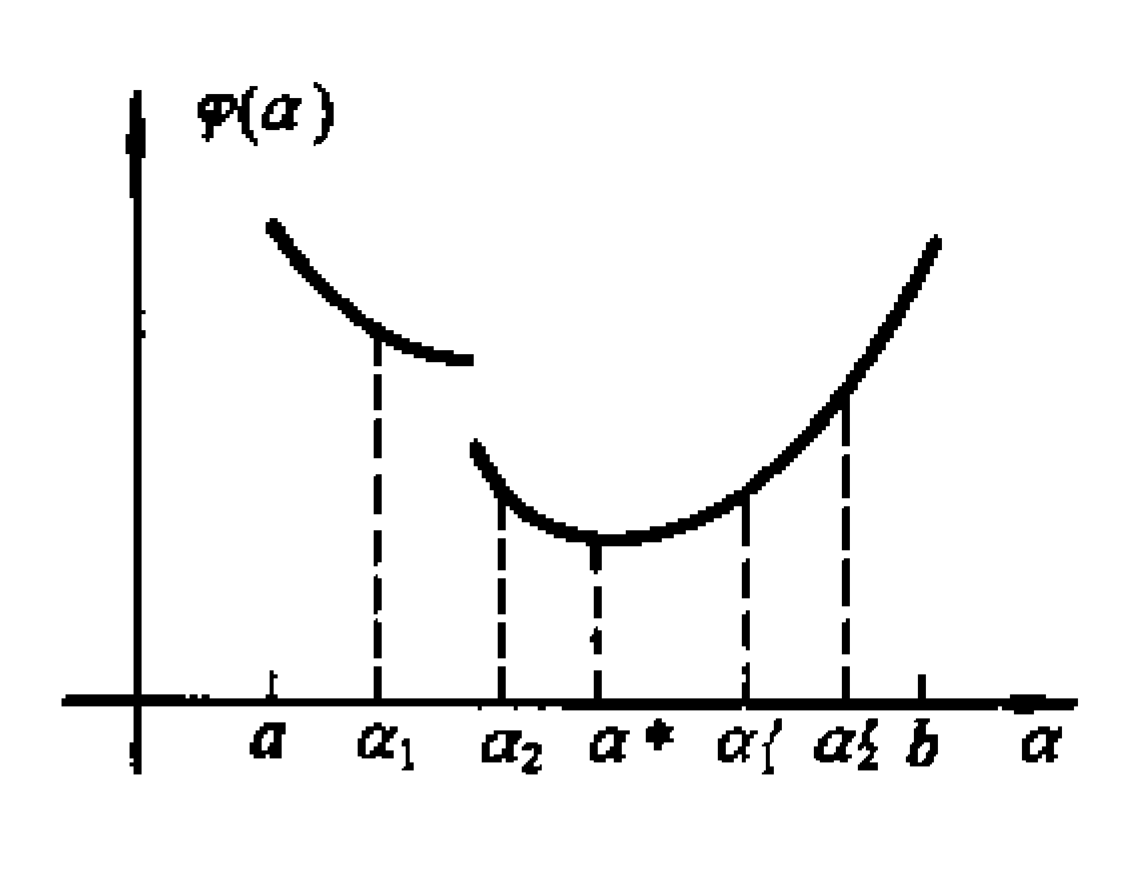


图 10.2 下单峰函数 $\varphi(\alpha)$

(i) 存在 $\alpha^* \in [a,b]$,使

$$\varphi(\alpha^*) = \min_{\alpha\in[a,b]} \varphi(\alpha)$$

(ii) 对任意的 α_1,α_2且 $a\leqslant\alpha_1<\alpha_2\leqslant b$,下面两个关系式成立:

$$\alpha_2 \leqslant \alpha^* \Leftrightarrow \varphi(\alpha_1) > \varphi(\alpha_2)$$

$$\alpha^* \leqslant \alpha_1 \Leftrightarrow \varphi(\alpha_1) < \varphi(\alpha_2)$$

设 $\varphi(\alpha)$为$[a,b]$上的下单峰函数. 为寻找 α^*,取 α_1,α_2: $a<\alpha_1<\alpha_2<b$. 由定义10.12知,若 $\varphi(\alpha_1)\geqslant\varphi(\alpha_2)$,则 $\alpha^* \in [\alpha_1,b]$;否则就有 $\alpha^* \in [a,\alpha_2]$. 这就是说,计算区间中两个相异的点的函数值便可将搜索区间缩小. 设 $\varphi(\alpha_1)\leqslant\varphi(\alpha_2)$,于是选留$[a,\alpha_2]$. 为对$[a,\alpha_2]$重复以上过程,注意到 $\alpha_1\in(a,\alpha_2)$及 $\varphi(\alpha_1)$为已知,故只须在(a,α_2)中选一个异于 α_1的点即可. 这样重复下去可以找到 α^* 的一个近似. 容易看出,按照这种思想构造的算法依赖于取点的某种法则.

由于实际计算的过程不可能无限的,故预先规定选点的总数不超过 n 个. 那么最初的搜索区间的长度和最终剩下的区间长度之比,可以作为取点好坏的标准,这个值越大便说明取点的方式越好. 换一种说法就是:设 L_n 是这样的初始区间的长度,使得存在一

种取点方式，取 n 个点后能把 L_n 缩短到1；最优取点的方式应使 L_n 为最大.

如上定义的 L_n 组成一个集合 $\{L_n\}$. 我们来看它的上确界应满足什么条件. 设 $\{L_i\}$ 的上确界为 $F_i, i=0,1,\cdots,n$，由于不取点或只取一点无法进行比较，故

$$F_0 = F_1 = 1 \qquad (10.4)$$

令 $L_n=b-a$. 设最初的两个取点为 α_1,α_2：$a<\alpha_1<\alpha_2<b$. 那么余下还可以取 $n-2$个点. 极小点 α^* 的位置只有两种可能，或 $\alpha^*\in[a,\alpha_1]$，或 $\alpha^*\in[\alpha_1,b]$. 若为前者，那么余下还可以取 $n-2$个点，从而

$$\alpha_1 - a \leqslant F_{n-2}$$

若为后者，则除了余下可取的 $n-2$个点外，还有 $\alpha_2\in(\alpha_1,b)$，故实际上等于可取 $n-1$个点，从而还有

$$b - \alpha_1 \leqslant F_{n-1}$$

将上述两式相加得到

$$L_n = b - a \leqslant F_{n-2} + F_{n-1}$$

由 $L_n\in\{L_n\}$的任意性可知

$$F_n \leqslant F_{n-2} + F_{n-1} \qquad (10.5)$$

事实上极易证明

$$F_n = F_{n-2} + F_{n-1} \quad (n \geqslant 2) \qquad (10.6)$$

定义10.13 按递推关系式(10.4)和(10.6)所得的数列称为 Fibonacci 数.

下面给出最初的几个 Fibonacci 数：

n	0	1	2	3	4	5	6	7	8	9
F_n	1	1	2	3	5	8	13	21	34	55

由前面的结论可以看出，取满足 $F_N\geqslant(b-a)/\varepsilon$ 的 N，经过 N 次取点后，最终的区间长度不超过 ε. 于是有

算法10.14(Fibonacci 法)

设 $\varphi(\alpha)$是$[a,b]$上的下单峰函数. 预先给定允许误差 $\varepsilon>0$和

能分别出函数值的最小的 $\delta>0$.

步0　求使 $F_n \geqslant (b-a)/\varepsilon$ 的最小的 n，记为 N；

步1　$k:=N$，并按下两式计算$[a,b]$的两个内点：

$$\alpha_l = a + \frac{F_{k-2}}{F_k}(b-a) \tag{10.7}$$

$$\alpha_r = a + \frac{F_{k-1}}{F_k}(b-a) \tag{10.8}$$

步2　$k:=k-1$；

步3　若 $\varphi(\alpha_l) \geqslant \varphi(\alpha_r)$ 则 $a:=\alpha_l$，$\alpha_l:=\alpha_r$，当 $k<2$ 时，停；当 $k>2$ 时，按(10.8)计算 α_r；当 $k=2$ 时，置 $\alpha_r=\alpha_l+\delta$，转步2. 否则，$b:=\alpha_r$，$\alpha_r:=\alpha_l$，当 $k<2$ 时，停；当 $k>2$ 时，按(10.7)计算 α_l；当 $k=2$ 时，置 $\alpha_l=\alpha_r-\delta$，转步2.

现在，我们来计算 F_{k-1}/F_k 的极限. 令

$$F_n - \alpha F_{n-1} = \beta(F_{n-1} - \alpha F_{n-2}) \tag{10.9}$$

$$F_n - \beta F_{n-1} = \alpha(F_{n-1} - \beta F_{n-2}) \tag{10.10}$$

$$\alpha + \beta = 1, \alpha \cdot \beta = -1 \tag{10.11}$$

则 Fibonacci 数之定义式(10.6)与(10.9)～(10.11)是等价的. 注意到 $F_0=F_1=1$ 以及利用(10.11)，易由(10.9)和(10.10)递推得

$$F_n - \alpha F_{n-1} = \beta^{n-1}(1-\alpha) = \beta^n \tag{10.12}$$

$$F_n - \beta F_{n-1} = \alpha^{n-1}(1-\beta) = \alpha^n \tag{10.13}$$

将上两式相减得

$$F_{n-1} = \frac{\beta^n - \alpha^n}{\beta - \alpha} \tag{10.14}$$

再由(10.11)解得 $\alpha=(1-\sqrt{5})/2$，$\beta=(1+\sqrt{5})/2$，代入(10.14)得

$$F_n = \frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^{n+1} - \left(\frac{1-\sqrt{5}}{2}\right)^{n+1}\right] \quad (n \geqslant 0) \tag{10.15}$$

最后，由(10.15)又得

$$\frac{F_{n-1}}{F_n} \to \frac{\sqrt{5}-1}{2} \triangleq \tau \approx 0.618$$

在(10.8)式和(10.7)式中分别以 τ 和$1-\tau$ 代替 F_{k-1}/F_k 和 F_{k-2}/F_k,将该算法稍加改动便得黄金分割法.

算法10.15(黄金分割法)

设 $\varphi(\alpha)$为$[a,b]$上的下单峰函数.取定容许误差 $\varepsilon>0$.

步1 用下列公式计算试探点

$$\alpha_l = a + (1-\tau)(b-a) \tag{10.16}$$

$$\alpha_r = a + \tau(b-a) \tag{10.17}$$

步2 若 $\varphi(\alpha_r)>\varphi(\alpha_l)$,则 $b:=\alpha_r,\alpha_r:=\alpha_l$,并由(10.16)计算新点 α_l;否则,$a:=\alpha_l,\alpha_l:=\alpha_r$,并由(10.17)计算新点 α_r;

步3 若 $b-a<\varepsilon$,停;否则转步2.

在具体使用公式(10.16)和(10.17)时,我们常常取 τ 的近似值0.618.对应的方法亦称为0.618法.因为黄金分割法(或0.618法)实际上是每次将区间长度用比例因子 τ(或0.618)去缩短之.所以可以断言这个算法的收敛速度不会优于线性敛速.方法的好处是此法简单且适用范围广.

2. 多项式逼近法

上一段的方法是在每一步将 φ 的两个函数值作比较.我们希望能够较充分地利用函数所给的信息,从而构造更为有效的方法.这导致了多项式逼近法.

三点二次插值

若已知$[a,b]$上的连续函数 $\varphi(\alpha)$在三点 $\alpha_1,\alpha_2,\alpha_3$的函数值分别为 $\varphi_1,\varphi_2,\varphi_3$.我们通过点$(\alpha_1,\varphi_1)$,$(\alpha_2,\varphi_2)$,$(\alpha_3,\varphi_3)$做二次插值多项式 $\hat{\varphi}(\alpha)$.然后用 $\hat{\varphi}(\alpha)$的极小点 $\hat{\alpha}$ 近似地作为 $\varphi(\alpha)$的极小点.这是三点二次插值的基本思想.

为保证二次式 $\hat{\varphi}(\alpha)$有极小点,条件"$\alpha_1<\alpha_2<\alpha_3,\varphi_1\geqslant\varphi_2,\varphi_2\leqslant\varphi_3$且后两不等式的等号不能同时成立"是充分条件.这样的(α_1,φ_1),(α_2,φ_2),(α_3,φ_3)称为一个三点模型.利用 Lagrange 插值公式于这三点得

$$\hat{\varphi}(\alpha) = \sum_{i=1}^{3} \frac{\prod_{j\neq i}(\alpha - \alpha_j)}{\prod_{j\neq i}(\alpha_i - \alpha_j)} \varphi_i \qquad (10.18)$$

令 $\hat{\varphi}'(\alpha)=0$，从而解得 $\hat{\alpha}\in(\alpha_1,\alpha_3)$：

$$\hat{\alpha}=\frac{1}{2}\frac{[(\alpha_2)^2-(\alpha_3)^2]\varphi_1+[(\alpha_3)^2-(\alpha_1)^2]\varphi_2+[(\alpha_1)^2-(\alpha_2)^2]\varphi_3}{(\alpha_2-\alpha_3)\varphi_1+(\alpha_3-\alpha_1)\varphi_2+(\alpha_1-\alpha_2)\varphi_3}$$

(10.19)

为作进一步的迭代，应该在(α_1,φ_1)，(α_2,φ_2)，(α_3,φ_3)以及新点$(\hat{\alpha},\varphi(\hat{\alpha}))$这四点中选出一个新的三点模型来. 新的三点模型应包括新点，并能使区间$[\alpha_1,\alpha_3]$缩短. 如果在这四点中找不出一个新的三点模型来(比如说，恰好有 $\hat{\alpha}=\alpha_2$)，则应先将 $\hat{\alpha}$ 在(α_1,α_3)中进行摄动得 $\hat{\hat{\alpha}}\in(\alpha_1,\alpha_3)$，使得可从$(\hat{\hat{\alpha}},\varphi(\hat{\hat{\alpha}}))$，$(\alpha_1,\varphi_1)$，$(\alpha_2,\varphi_2)$以及$(\alpha_3,\varphi_3)$中选出一个新三点模型来.

二点三次插值

若已知$[a,b]$上的连续函数$\varphi(\alpha)$在 α_1，α_2处的函数值为 φ_1，φ_2和导数值 φ_1'，φ_2'，则可以作一个三次多项式

$$\hat{\varphi}(\alpha) = c_0 + c_1(\alpha - \alpha_1) + c_2(\alpha - \alpha_1)^2 + c_3(\alpha - \alpha_1)^3$$

(10.20)

其中，系数 c_0，c_1，c_2，c_3可从下面的线性方程组中解出：

$$\begin{cases} \varphi_1 = \hat{\varphi}(\alpha_1) = c_0 \\ \varphi_1' = \hat{\varphi}'(\alpha_1) = c_1 \\ \varphi_2 = \hat{\varphi}(\alpha_2) = c_0 + c_1(\alpha_2 - \alpha_1) + c_2(\alpha_1 - \alpha_1)^2 + c_3(\alpha_2 - \alpha_1)^3 \\ \varphi_2' = \hat{\varphi}'(\alpha_2) = c_1 + 2c_2(\alpha_2 - \alpha_1) + 3c_3(\alpha_2 - \alpha_1)^2 \end{cases}$$

(10.21)

为了使 $\hat{\varphi}(\alpha)$在(α_1,α_2)中有一个极小点，条件“$\alpha_1<\alpha_2$，$\varphi_1'<0$且$\varphi_2'>0$”是一个充分条件. 在此条件下，对由(10.21)定义的(10.20)作详细的推导，可以解出 $\hat{\varphi}(\alpha)$的极小点 $\hat{\alpha}\in(\alpha_1,\alpha_2)$为

$$\hat{\alpha} = \alpha_2 - (\alpha_2 - \alpha_1)\frac{\varphi_2' + u_2 - u_1}{\varphi_2' - \varphi_1' + 2u_2} \qquad (10.22)$$

其中

$$u_1 = \varphi_2' + \varphi_1' - 3(\varphi_2 - \varphi_1)/(\alpha_2 - \alpha_1)$$
$$u_2 = [(u_1)^2 - \varphi_1' \cdot \varphi_2']^{1/2}$$

新的点 $\hat{\alpha}$ 就近似地看作 $\varphi(\alpha)$ 的极小点. 余下的做法和三点二次插值法是类似的. 故从略.

收敛性

一般认为,只知道函数值以用三点二次插值为宜. 如果还能利用导数值则以二点三次插值为宜. 更高次的代数插值会使计算来得复杂. 如果已知所给的函数在初始区间 $[a,b]$ 上是下单峰的并且是可导的,则能够证明这两种方法都是收敛的. 在一般的情况下,要对算法作一些修改,使得修改后的算法具全局收敛性. 全局收敛性定理可为修改作参考. 在这方面已有大量工作,可参看有关的书籍,例如文献[11]和[23].

收敛速度

已经证明,三点二次插值法的收敛阶为1.3,二点三次插值法的收敛阶为2. 证明的方法是先对(10.19)、(10.22)作详细的分析(假定 $\varphi(\alpha)$ 足够光滑),从而得出一个差分方程,然后对此差分方程求解以获得收敛的阶. 下面,我们介绍一种简单的(然而是粗糙的)收敛阶的估计方法. 用它可以对任意构造的插值多项式法预先进行收敛速度的估计. 以二点三次插值为例,先注意到在(10.22)中按迭代法,可分别视 $\alpha_1, \alpha_2, \hat{\alpha}$ 为 $\alpha_{k-1}, \alpha_k, \alpha_{k+1}$.

首先,由插值多项式理论,设 $\varphi \in C^4$,则有

$$\varphi(\alpha) = P_3(\alpha) + \frac{\varphi^{(4)}(\xi)}{4!}(\alpha - \alpha_k)^2(\alpha - \alpha_{k-1})^2 \quad (10.23)$$

其中 ξ 在 $\alpha, \alpha_{k-1}, \alpha_k$ 之间, $P_3(\alpha) = \hat{\varphi}(\alpha)$ 如(10.20)、(10.21)所定义.

设 α^* 为 $\varphi(\alpha)$ 的极小点,则 $\varphi'(\alpha^*) = 0$. 对(10.23)先关于 α 求导,然后将 α^* 代入得到

$$P_3'(\alpha^*) = \frac{\varphi^{(4)}(\xi)}{12}\{(\alpha_{k-1} - \alpha^*)^2(\alpha_k - \alpha^*) + (\alpha_{k-1} - \alpha^*)(\alpha_k - \alpha^*)^2\}$$
$$+ \left[\frac{\varphi^{(4)}(\xi)}{4!}\right]'_{\alpha=\alpha^*}(\alpha_k - \alpha^*)^2(\alpha_{k-1} - \alpha^*)^2 \quad (10.24)$$

又,按插值求极小的方法有 $P_3'(\alpha_{k+1}) = 0$. 在(10.24)中记

$$\varepsilon_{k-1} = \alpha_{k-1} - \alpha^*, \varepsilon_k = \alpha_k - \alpha^*$$

并假定$[\varphi^{(4)}(\xi)]'_{\alpha=\alpha^*}$和$[\varphi^{(4)}(\xi)]$是有界的，还要假定$|\varepsilon_k| < |\varepsilon_{k-1}|$（注意，这些都是通常证明中的假定），则当$\{\alpha_k\}$与$\alpha^*$充分接近时，由(10.24)近似地得到

$$P'_3(\alpha_{k+1}) - P'_3(\alpha^*) = -\frac{\varphi^{(4)}(\xi)}{12}\varepsilon_k\varepsilon_{k-1}^2 \tag{10.25}$$

对(10.25)左端利用一次中值定理，有

$$P''_3(\eta_k)\varepsilon_{k+1} = -\frac{\varphi^{(4)}(\xi)}{12}\varepsilon_k\varepsilon_{k-1}^2 \tag{10.26}$$

其中$\varepsilon_{k+1} = \alpha_{k+1} - \alpha^*$，$\eta_k$在$\alpha_{k+1}$与$\alpha^*$之间.

最后，当$\alpha_{k-1}, \alpha_k, \alpha_{k+1}$充分接近于$\alpha^*$时，我们得到

$$\varepsilon_{k+1} \doteq M\varepsilon_k\varepsilon_{k-1}^2 \tag{10.27}$$

其中$M = -\frac{1}{12}\varphi^{(4)}(\alpha^*)/\varphi''(\alpha^*)$.

令$y_i = \ln\{|M|^{1/2}|\varepsilon_i|\}$，$i = k-1, k, k+1$，由(10.27)得如下的差分方程：

$$y_{k+1} = y_k + 2y_{k-1} \tag{10.28}$$

它的特征方程为

$$\lambda^2 - \lambda - 2 = 0 \tag{10.29}$$

上方程的最大正根为2，故应有

$$y_{k+1} - 2y_k \to 0, k \to \infty \tag{10.30}$$

由(10.30)立即可推得欲证之式

$$\frac{|\alpha_{k+1} - \alpha^*|}{|\alpha_k - \alpha^*|^2} = \frac{|\varepsilon_{k+1}|}{\varepsilon_k^2} \longrightarrow 1 \tag{10.31}$$

§10.3 R^n中的最优化

今考虑R^n中的无约束最优化问题$\min\limits_{x\in R^n} f(x)$. 设已给定$x^0 \in R^n$且$x^0$非该问题的解. 如何找一个$x^1 \in R^n$，使得$f(x^1) < f(x^0)$？一个自然的想法是从$x^0$处找一个下降方向$d^0$，即这样的$d^0$，使得对充分小的$\alpha > 0$恒有

$$f(x^0 + \alpha d^0) < f(x^0) \tag{10.32}$$

为使方法更有效，应取 $x^1=x^0+\alpha_0 d^0$，使得 $\alpha_0>0$满足

$$f(x^0 + \alpha_0 d^0) = \min_{\alpha\geqslant 0} f(x^0 + \alpha d^0) \tag{10.33}$$

若记 $\varphi(\alpha)=f(x^0+\alpha d^0)$，则(10.33)可以描述为：寻找 $\alpha_0\geqslant 0$，使得 $\varphi(\alpha_0)=\min\limits_{\alpha\geqslant 0}\varphi(\alpha)$. 这就是前节所述的线搜索. 因此可以说，对给定的问题$\min\limits_{x\in R^n} f(x)$和给定的点 x，如何寻找 x 处的“最好的”下降方向，便是 R^n 中的最优化的一个重要课题了.

1. 最速下降法

早在1847年，Cauchy 就注意到 $f(x)$在 x 处的最速下降方向为$-\nabla f(x)$，即 $f(x)$的负梯度方向. 这个推导是很简单的. 设 $f\in C^1$，则

$$f(x + \alpha d) \doteq f(x) + \alpha\nabla f(x)^T d \tag{10.34}$$

注意到

$$\nabla f(x)^T d = \|\nabla f(x)\| \cdot \|d\| \cdot \cos\langle f(x), d\rangle$$

若$\nabla f(x)\neq 0$，则 $d=-\nabla f(x)/\|\nabla f(x)\|$就是问题

$$\min_{\|d\|\leqslant 1} \nabla f(x)^T d \tag{10.35}$$

的解. 即$-\nabla f(x)$为 $f(x)$在 x 处的最速下降方向.

算法10.16(最速下降法)

设 $f: R^n\to R, f\in C^1$.

步0　任取初始点 $x^0\in R^n$，$k:=0$；

步1　计算$\nabla f(x^k)$，若$\nabla f(x^k)=0$，停. 否则

步2　$d^k:=-\nabla f(x^k)$，求 $\alpha_k\geqslant 0$：

$$f(x^k+\alpha_k d^k)=\min_{\alpha\geqslant 0} f(x^k+\alpha d^k) \tag{10.36}$$

$x^{k+1}:=x^k+\alpha_k d^k$，$k:=k+1$，返步1.

容易看出，这里的步2就是算法10.3的步1. 两厢对照可知

$$A = SG \tag{10.37}$$

其中 G 表示定方向的映射，S 表示线搜索. 注意，满足(10.36)的 α_k 可能不止一个. 故 S 应理解为点到集的映射. 从而 $A=SG$ 也是点

到集的映射.

为证由(10.37)所定义的映射 A 是闭的,我们有

定义10.17 $S:R^{2n}\to R^n$,其定义为

$$S(x,d)=\{y=x+\bar{\alpha}d:\bar{\alpha}\geqslant 0 \text{ 且 } f(y)=\min_{\alpha\geqslant 0}f(x+\alpha d)\}$$

定理10.18 设 f 在 R^n 上连续且 $d\neq 0$,则 $S(x,d)$ 在 (x,d) 处是闭的.

证 设 $x^k\to x,d^k\to d\neq 0,y^k\to y,y^k\in S(x^k,d^k)$. 欲证

$$y\in S(x,d)$$

首先,由 $y^k\in S(x^k,d^k)$ 知,存在 $\alpha_k\geqslant 0$,使得

$$y^k=x^k+\alpha_k d^k,\alpha_k\geqslant 0 \tag{10.38}$$

$$f(y^k)\leqslant f(x^k+\alpha d^k),\alpha\in[0,\infty) \tag{10.39}$$

由(10.38)得 $\alpha_k=\|y^k-x^k\|/\|d^k\|\to\|y-x\|/\|d\|\triangleq\bar{\alpha}\geqslant 0$(由 $d\neq 0$).
故在(10.38)中令 $k\to\infty$ 得

$$y=x+\bar{\alpha}d,\bar{\alpha}\geqslant 0 \tag{10.40}$$

再在(10.39)中固定 α,令 $k\to\infty$ 得

$$f(y)\leqslant f(x+\alpha d),\alpha\in[0,+\infty) \tag{10.41}$$

上式表明 $y\in S(x,d)$. 证毕.

现在,我们可以给出

定理10.19(最速下降法的收敛性) 设 $f:R^n\to R,f\in C^1$. 若算法10.16终止于某 x^k,则 $\nabla f(x^k)=0$. 又若算法产生一个无穷点列 $\{x^k\}$ 且 $\{x^k\}$ 有界,则 $\{x^k\}$ 的任一聚点 $\bar{x}$ 都满足 $\nabla f(\bar{x})=0$.

证 前一部分已由算法保证. 故只须证后一部分.

首先,定义解集合

$$\Gamma=\{x\in R^n:\nabla f(x)=0\}$$

定义

$$G:R^n\to R^{2n},G(x)=(x,-\nabla f(x))$$

$$S:R^{2n}\to R^n,S=S(x,-\nabla f(x))$$

由 G 是连续的点到点的映射,S 在点 $(x;-\nabla f(x))$ $(\nabla f(x)\neq 0)$ 处是闭的,由推论10.10知 $A=SG$ 在 Γ 外是闭的.

其次，对 $x \in \Gamma$，即 $\nabla f(x) \neq 0$，对小的 $\alpha > 0$ 有

$$f(x - \alpha \nabla f(x)) = f(x) - \alpha \|\nabla f(x)\|^2 + o(\alpha) < f(x) \tag{10.42}$$

故取下降函数(关于如上定义的 Γ 和 A)为 $\omega(x) = f(x)$.

利用全局收敛性定理(定理10.6)可知结论为真. 证毕.

一般地说，利用 §10.2提供的线搜索法求出(全局)极小点是不可能的也是不经济的. 这就导致了下面的

定理10.20 下面的三种线搜索(其中第二种要求 $d \neq 0$)在 (x,d)处都是闭的，其中 $f: R^n \to R$ 为连续.

类型 I 取定 $\delta > 0$,

$$S^\delta(x,d) = \{y = x + \bar{\alpha} d : f(y) = \min_{0 \leqslant \alpha \leqslant \delta} f(x + \alpha d)\}$$

类型 II 取定 $\varepsilon > 0$,

$$^\varepsilon S(x,d) = \{y = x + \bar{\alpha} d : f(y) \leqslant \min_{\alpha \geqslant 0} f(x + \alpha d) + \varepsilon\}$$

类型 III 取定 $\varepsilon > 0, \delta > 0$,

$$^\varepsilon S^\delta(x,d) = \{y = x + \bar{\alpha} d : f(y) \leqslant \min_{0 \leqslant \alpha \leqslant \delta} f(x + \alpha d) + \varepsilon\}$$

类型 I 表明极小问题总限定在区间$[0,\delta]$上，而不是在半直线$[0,\infty)$上. 类型 II 表明求极小值时可以允许一个误差. 类型 III 是前两者的结合. 证明方法与定理10.18的证明类似.

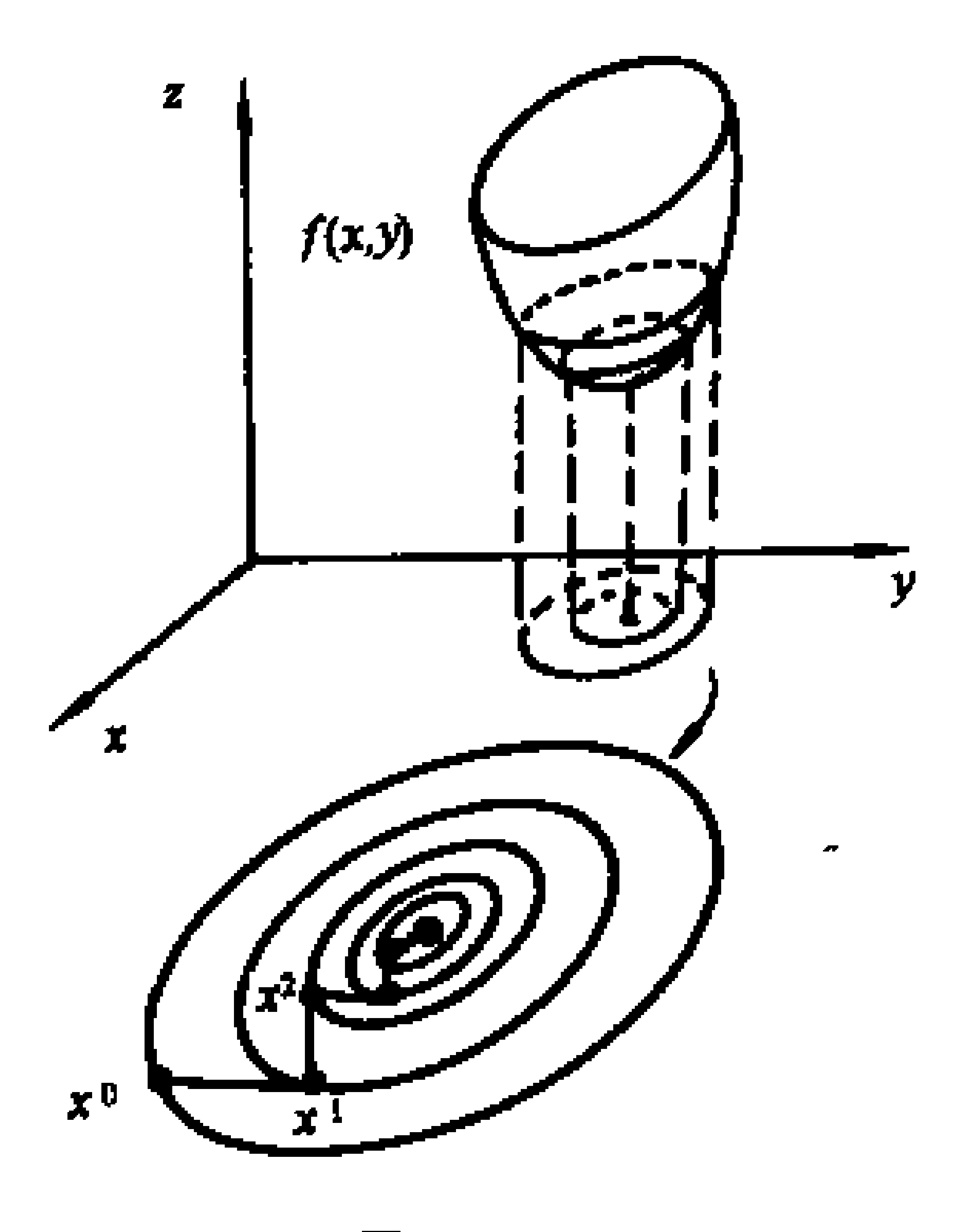


图 10.3

顾名思义，最速下降法应该是一种很好的方法，其实不然. 由

$$\frac{\mathrm{d}}{\mathrm{d}\alpha} f(x^k + \alpha d^k)|_{\alpha = \alpha_k} = 0$$

得

$$\nabla f(x^k + \alpha_k d^k)^T d^k = - \nabla f(x^{k+1})^T \nabla f(x^k) = 0$$

这表明由最速下降法产生的两个

相邻的方向是互相垂直的. 参照图10.3可知，最速下降法在求解问题的最初几步的进程比较快，到了后来就越来越慢. 有例表明它的收敛速度只能是线性的(参见习题6).

定理10.21 设 $f:R^n\to R$ 在开凸集 D_0上二次连续可微，设存在 $M\geqslant m>0$，使得在 D_0上有

$$m\|d\|^2\leqslant d^T\nabla^2 f(x)d\leqslant M\|d\|^2,\forall\ d\in R^n \tag{10.43}$$

再设给定的初始点 x^0使得水平集

$$X_0=\{x\in R^n;f(x)\leqslant f(x^0)\}\subseteq D_0 \tag{10.44}$$

且 X_0有界. 若最速下降法(算法10.16)产生无穷点列 $\{x^k\}$，则

(i) $x^k\to x^*$，且 x^* 为 $\min\limits_{x\in R^n} f(x)$ 的唯一的全局极小点；

(ii) 成立着以下的收敛速度估计

$$\|x^k-x^*\|\leqslant c\, q^{k/2},c<+\infty \tag{10.45}$$

$$f(x^k)-f(x^*)\leqslant q^k(f(x^0)-f(x^*)) \tag{10.46}$$

其中 $q=1-(m/M)^2<1$.

证 (i) 由下降法知 $\{x^k\}\subseteq X_0$；由 X_0为有界闭集知 $\{x^k\}$有一聚点在 X_0中，设为 x^*；由定理10.19知 $\nabla f(x^*)=0$；由 $X_0\subseteq D_0$知 $x^*\in D_0$；由(10.43)知 f 在 D_0上为严格凸；从而 $\nabla f(x^*)=0$便意味着 x^* 为 f 在 D_0上的唯一全局极小点. 复由 $X_0\subseteq D_0$知 x^* 还是 f 在 R^n 上的唯一全局极小点. 由唯一性及 $\{x^k\}\subseteq X_0$的有界性即推得 $x^k\to x^*$.

(ii) 由 $\nabla f(x^*)=0$，$x^*\in D_0$及由 Taylor 公式，利用(10.43)知，对$\forall\ x\in D_0$，有

$$\frac{m}{2}\|x-x^*\|^2\leqslant f(x)-f(x^*)\leqslant\frac{M}{2}\|x-x^*\|^2 \tag{10.47}$$

另外，由

$$|(\nabla f(x)-\nabla f(x^*))^T(x-x^*)|\leqslant\|\nabla f(x)\|\|x-x^*\|$$

对 $\nabla f(x)-\nabla f(x^*)$利用中值定理，结合(10.43)得

$$\|\nabla f(x)\|\|x-x^*\|\geqslant m\|x-x^*\|^2$$

此即，对$\forall\ x\in D_0$有

$$\|x - x^*\| \leqslant \frac{1}{m}\|\nabla f(x)\| \tag{10.48}$$

结合(10.47)右端之不等式和(10.48)，得到

$$\frac{2m^2}{M}(f(x) - f(x^*)) \leqslant \|\nabla f(x)\|^2, \forall\, x \in D_0 \tag{10.49}$$

又因为

$$f(x)-f(x^k)=\nabla f(x^k)^T(x-x^k)$$
$$+\frac{1}{2}(x-x^k)^T\nabla^2 f(x^k+\theta(x-x^k))(x-x^k)$$

其中 $\theta\in(0,1)$，用 $x=x^k-\alpha\nabla f(x^k)$代入上式后再利用(10.43)右端之不等式，得

$$f(x^k - \alpha\nabla f(x^k)) - f(x^k) \leqslant - \alpha\left(1 - \frac{\alpha M}{2}\right)\|\nabla f(x^k)\|^2 \tag{10.50}$$

今在(10.50)左端取 $\alpha=\alpha_k$，在其右端取 $\alpha=1/M$(即关于 α 取最大)，不等式依然成立. 故

$$f(x^{k+1}) - f(x^k) \leqslant - \frac{1}{2M}\|\nabla f(x^k)\|^2 \tag{10.51}$$

今在(10.49)中取 $x=x^k$ 后，代入(10.51)之右端得

$$f(x^{k+1}) - f(x^k) \leqslant - \frac{m^2}{M^2}(f(x^k) - f(x^*))$$

从而解得

$$f(x^{k+1}) - f(x^*) \leqslant (1 - \frac{m^2}{M^2})(f(x^k) - f(x^*))$$
$$\triangleq q(f(x^k) - f(x^*)) \tag{10.52}$$

用(10.52)递推地得(10.46).

又由(10.47)左端之不等式及(10.46)得

$$\|x^k - x^*\|^2 \leqslant \frac{2}{m}(f(x^k) - f(x^*))$$
$$\leqslant \frac{2}{m}(f(x^0) - f(x^*))q^k \triangleq c^2q^k$$

此即(10.45)，证毕.

总之，我们已经比较系统地介绍了最速下降法．最速下降法虽有缺陷，但通过对它的研究可以得到启迪，导致产生更好的方法．

2. Newton 法简介

设 $f:R^n\to R, f\in C^2$，取定 $x^0\in R^n$．作 f 在 x^0处的二阶近似

$$\hat{f}(x)=f(x^0)+\nabla f(x^0)^T(x-x^0)+\frac{1}{2}(x-x^0)^T\nabla^2 f(x^0)(x-x^0)$$

设$\nabla^2 f(x^0)$为正定，则$\hat{f}(x)$的极小点即满足$\nabla \hat{f}(x)=0$的点，故有

$$\nabla f(x^0)+\nabla^2 f(x^0)(x-x^0)=0 \qquad (10.53)$$

把(10.53)的解近似地看作 $f(x)$的极小点，得

$$x^1=x^0-\nabla^2 f(x^0)^{-1}\nabla f(x^0)$$

从而构造迭代公式

$$x^{k+1}=x^k-\nabla^2 f(x^k)^{-1}\nabla f(x^k) \qquad (10.54)$$

利用(10.54)求解$\min\limits_{x\in R^n} f(x)$称为简单 Newton 法，此法也是一种多项式逼近法，它是利用了函数在某点的值、导数值及二阶导数值(Hesse 阵)的信息．

Newton 法的收敛阶为2，这一点极易证得．

定理10.22 设 $f\in C^3$，设在 x^*处有$\nabla f(x^*)=0$，$\nabla^2 f(x^*)$非异，则 Newton 法的收敛阶为2．

证 记

$$A(x)=x-\nabla^2 f(x)^{-1}\nabla f(x)$$

则(10.54)就是

$$x^{k+1}=A(x^k) \qquad (10.55)$$

由条件$\nabla f(x^*)=0$知 $x^*=A(x^*)$，从而由上式得

$$\begin{aligned} x^{k+1}-x^* &= A(x^k)-A(x^*) \\ &= A'(x^*)(x^k-x^*) \\ &\quad +\frac{1}{2}(x^k-x^*)^T A''(\bar{x})(x^k-x^*) \end{aligned} \qquad (10.56)$$

其中 $\bar{x}=x^*+\theta(x^k-x^*)$，$\theta\in(0,1)$．又

$$A'(x)|_{x=x^*} = I - \{[\nabla^2 f(x)^{-1}]' \nabla f(x) + \nabla^2 f(x)^{-1} \nabla^2 f(x)\}|_{x=x^*} = 0$$

故由(10.56)得

$$\|x^{k+1} - x^*\| = \frac{1}{2}\|(x^k - x^*)^T A''(\overline{x})(x^k - x^*)\| \leqslant \frac{1}{2}\|A''(\overline{x})\| \|x^k - x^*\|^2 \leqslant M\|x^k - x^*\|^2 \tag{10.57}$$

其中由 $f \in C^3$ 知 $\|A''(\overline{x})\| \leqslant 2M$ 对某个 M 成立，证毕.

简单的例子表明 Newton 法时某些问题不收敛，从而便产生了如下的修改型.

修改型 I 取 $d^k = -\nabla^2 f(x^k)^{-1} \nabla f(x^k)$，取 α_k：

$$f(x^k + \alpha_k d^k) = \min_{\alpha \geqslant 0} f(x^k + \alpha d^k) \tag{10.58}$$

但是 $\nabla^2 f(x^k)$ 可能奇异，于是又有

修改型 II 取

$$d^k = -[\varepsilon_k I + \nabla^2 f(x^k)]^{-1} \nabla f(x^k), \varepsilon_k > 0.$$

其中 $\varepsilon_k I + \nabla^2 f(x^k)$ 为正定，且当 $k \to \infty$ 时 $\varepsilon_k \to 0$. 取 α_k：

$$f(x^k + \alpha_k d^k) = \min_{\alpha \geqslant 0} f(x^k + \alpha d^k) \tag{10.59}$$

Newton 法的修改型和最速下降法可用下面的方式得到统一. 记 H_k 为 n 阶阵，取

$$d^k = -H_k \nabla f(x^k) \tag{10.60a}$$

$$x^{k+1} = x^k + \alpha_k d^k; f(x^{k+1}) = \min_{\alpha \geqslant 0} f(x^k + \alpha d^k) \tag{10.60b}$$

在(10.60)中取 $H_k \equiv I$ 得最速下降法；取 $H_k = \nabla^2 f(x^k)^{-1}$，或取 $H_k = [\varepsilon_k I + \nabla^2 f(x^k)]^{-1}$ 得修改型.

最后我们指出，对 $\nabla f(x^k) \neq 0$，为使(10.60b)得到的 x^{k+1} 满足 $f(x^{k+1}) < (f(x^k)$，简单的办法是在(10.60a)中取 H_k 为对称正定阵，请读者自证之.

习　题

1. 证明推论10.9和推论10.10.

2. 证明，如果 A 是一个连续的点到点的映射，那么在全局收敛性定理中可去掉假设(ii).

3. 证明定理10.20.

4. 设 c 给定且 $c\in[0,1]$，定义

$$F(x,d)=\left\{y=x+\alpha d: 0\leqslant\alpha<\infty,\ |\alpha-\bar{\alpha}|\leqslant c\bar{\alpha},\ \frac{\mathrm{d}}{\mathrm{d}\alpha}f(x+\bar{\alpha}d)=0\right\}$$

证明，若 $d\neq 0$ 且 $\frac{\mathrm{d}}{\mathrm{d}\alpha}f(x+\alpha d)$ 是连续的，则 F 在 (x,d) 处是闭的.

5. 证明，若在(10.60a)中取 H_k 为对称正定阵，则若 $\nabla f(x^k)\neq 0$，由(10.60b)得到的 x^{k+1} 满足 $f(x^{k+1})<f(x^k)$.

6. 求函数 $f(x_1,x_2)=\frac{1}{3}(x_1)^2+\frac{1}{2}(x_2)^2$ 的极小点，取初始点为 $(3,2)^T$；再利用最速下降法进行计算，并给出收敛速度.

7. 考虑修正的 Newton 法对 $f\in C^1$ 的如下的算法：

$$x^{k+1}=x^k-\alpha_k H_k\nabla f(x^k)$$

其中 H_k 为 n 阶阵，α_k 是问题

$$\min_{\alpha\geqslant 0} f(x^k-\alpha H_k\nabla f(x^k))$$

的一个解. 令

$$\mathscr{H}=\left\{H\in R^{n\times n}: \begin{array}{l}H\text{ 为对称，对给定的 } b\geqslant a>0,\\ H\text{ 的所有特征值介于 } a、b\text{ 之间}\end{array}\right\}$$

令 H_k 取自 $\mathscr{H}$ 中的任一元. 假定算法所产生的点列 $\{x^k\}$ 是有界的，证明 $\{x^k\}$ 的任一聚点 x^* 都满足 $\nabla f(x^*)=0$.

8. 考虑另一种修正算法，也用于 $f\in C^1$，

$$x^{k+1}=x^k-\alpha_k B(x^k)\nabla f(x^k)$$

其中 $B(x)$ 是 n 阶对称正定阵，且它是 x 的连续函数，α_k 满足

$$f(x^{k+1})=\min_{\alpha\geqslant 0} f(x^k-\alpha B(x^k)\nabla f(x^k))$$

假定由这个算法所产生的点列 $\{x^k\}$ 是有界的，证明 $\{x^k\}$ 的任一聚点 x^* 满足 $\nabla f(x^*)=0$.

9. 对于 $\min\limits_{x\in R^n} f(x)$，$f\in C^1$，有人提出如下的方法，称为平行切线法(或写为 Partan 法)：

给定 x^0，令 $d^1=-\nabla f(x^0)$，沿 d^1 极小化 f 得 x^1；令 $d^2=-\nabla f(x^1)$，以

x^1为出发点，沿 d^2极小化 f 得 x^2(即二步最速下降法)；再令 $d^3=x^2-x^0$(称为加速方向)，再以 x^0为出发点，沿 d^3极小化 f 得 x^3. 这样便完成了方法的初始工作. 接下去是交替地沿负梯度方向和加速方向作线搜索(它的第一步如图10.4中虚线所示)，线搜索的方向总结如下：

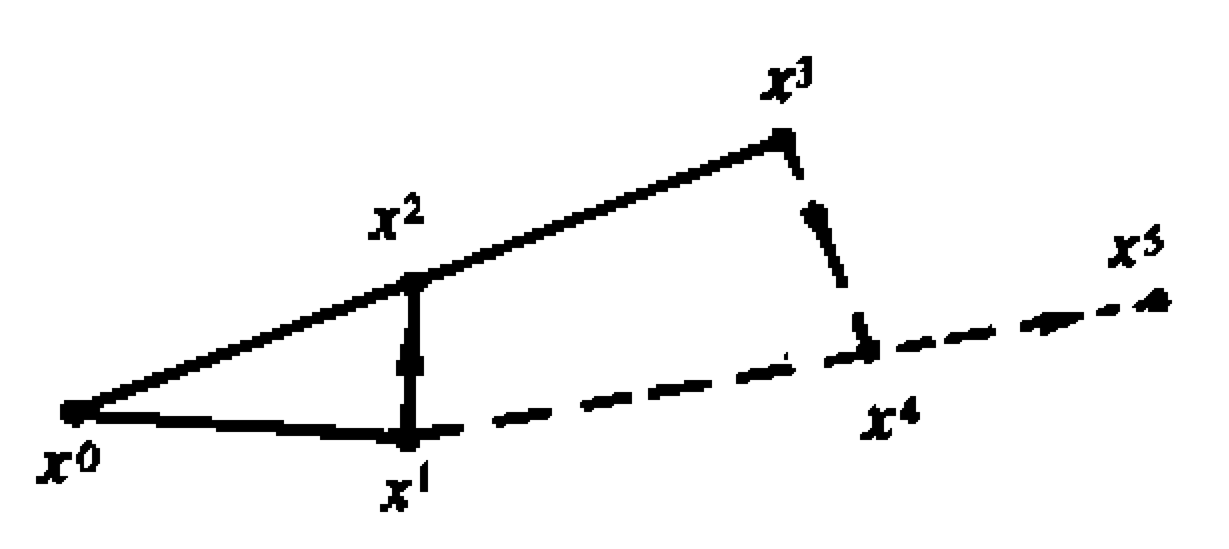


图 10.4

$$d^k = -\nabla f(x^{k-1}),\ k = 1,2,4,6,\cdots$$

$$d^k = x^{k-1} - x^{k-3},k = 3,5,7,9,\cdots$$

试用这个方法对第6题的问题做三、五步，看看结果如何.

第十一章　共轭法和拟 Newton 法

本章所介绍的方法都可以看作介于最速下降法和 Newton 法之间的方法. 一方面要克服最速下降法的慢收敛，另一方面要避免 Newton 法中计算 Hesse 阵的逆的烦恼. 由于二次可微函数在极小点附近呈二次性态，因此本章研究问题的出发点是导出的算法用于二次问题都能在有限步内得解. 具有这个性质的算法也称其有二次终结性.

所谓二次问题即

$$\min_{x\in R^n} f(x) = \frac{1}{2}x^TQx - b^Tx \tag{11.1}$$

其中 Q 为 n 阶对称正定阵，$b\in R^n$.

§11.1　共轭方向法

我们从叙述下面的概念开始.

定义 11.1　设 Q 为 n 阶对称正定阵，设 $d^1,d^2\in R^n$. 若 $(d^1)^TQd^2=0$，则称 d^1,d^2 关于 Q 是互为共轭的. 又若 $d^1,d^2,\cdots$ 关于 Q 两两共轭，则称它们是一组 Q 共轭向量.

令 $Q=I$，则共轭的概念化为正交的概念. 因此，Q 共轭也称为 Q 正交.

定理11.2　若非零向量集 $d^0,d^1,\cdots,d^k(k\leqslant n-1)$ 是一组 Q 共轭向量，则它们是线性无关的.

证　设有 $\alpha_i,i=0,1,\cdots,k$，使得

$$\alpha_0d^0+\alpha_1d^1+\cdots+\alpha_kd^k=0$$

分别以 $(d^i)^TQ,i=0,1,\cdots,k$ 左乘上式，得

$$\alpha_i(d^i)^TQd^i=0, i=0,1,\cdots,k \tag{11.2}$$

由 Q 为正定且 $d^i\neq 0, i=0,1,\cdots,k$，得 $\alpha_i=0$. 证毕.

设给定了对称正定阵 Q 及 n 个线性无关的向量. 用 Gram-Schmidt 正交化方法于这 n 个向量，从而得到一组 Q 共轭向量. 在习题中，我们给出求 Q 共轭向量的两种方法.

设已给定 n 个非零的 Q 共轭向量 $d^0,d^1,\cdots,d^{n-1}$. 求解二次问题(11.1)的共轭方向法可描述如下：任意给定 x^0，在 $x^0+\alpha d^0$ 上求 f 的极小得 x^1；再在 $x^1+\alpha d^1$ 上求 f 的极小得 x^2，…，最后，在 $x^{n-1}+\alpha d^{n-1}$ 上求 f 的极小得 x^n. x^n 便是(11.1)的解.

例11.3 设

$$Q=\begin{pmatrix}2 & 1\\ 1 & 4\end{pmatrix}, b=0$$

则

$$f(x)=\frac{1}{2}x^TQx=(x_1)^2+x_1x_2+2(x_2)^2$$

易知它的极小点为

$$x^*=(0,0)^T$$

选二个 Q 共轭向量

$$d^0=(1,0)^T, d^1=(-1,2)^T$$

取 $x^0=(2,-3)^T$，则

$$f(x^0+\alpha d^0)=(2+\alpha)^2+(2+\alpha)(-3)+2\cdot(-3)^2$$

易解得 $\alpha_0=-1/2$，从而 $x^1=x^0+\alpha_0d^0=(3/2,-3)^T$；同理得 $\alpha_1=3/2$，以及 $x^2=x^1+\alpha_1d^1=(0,0)^T$.（参见图11.1.）

下面对方法进行分析. 设 Q 为对称正定，取定 $x^k,d^k\in R^n$，则

$$f(x^k+\alpha d^k)=\frac{1}{2}(x^k+\alpha d^k)^TQ(x^k+\alpha d^k)-b^T(x^k+\alpha d^k) \tag{11.3}$$

为求极小，令

$$0=\frac{\mathrm{d}}{\mathrm{d}\alpha}f(x^k+\alpha d^k)=(d^k)^TQ(x^k+\alpha d^k)-b^Td^k \tag{11.4}$$

解上式得

$$\alpha_k=-\frac{(d^k)^T(Qx^k-b)}{(d^k)^TQd^k}=-\frac{(d^k)^Tg^k}{(d^k)^TQd^k} \tag{11.5}$$

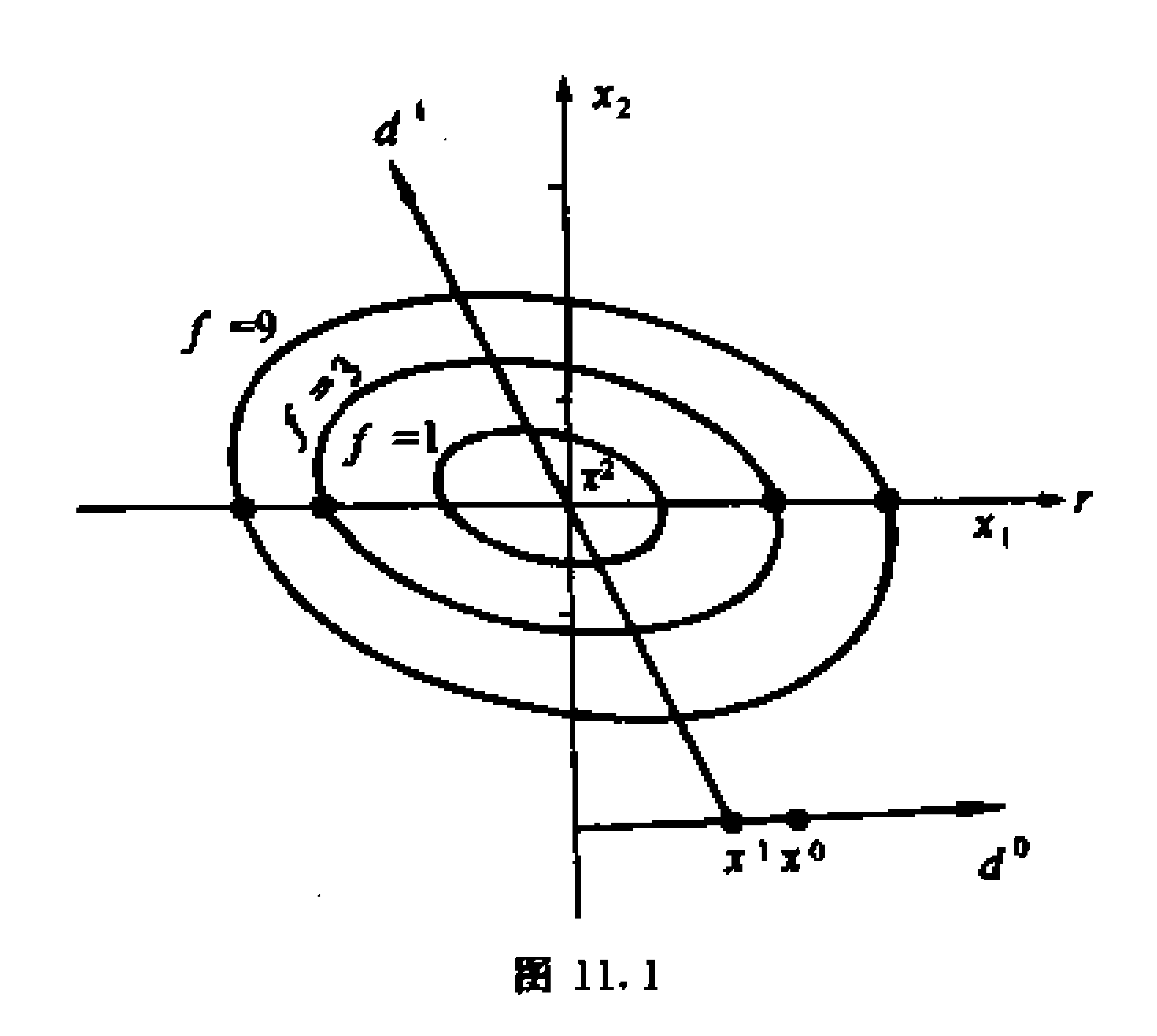


图 11.1

其中

$$g^k = Qx^k - b = \nabla f(x^k) \tag{11.6}$$

注意，在 Q 为对称正定时，x^* 为问题(11.1)的解的充要条件是 $\nabla f(x^*)=0$. 即

$$Qx^* = b \tag{11.7}$$

记 B_k 为由 $d^0, d^1, \cdots, d^{k-1}$ 生成的子空间. 记 x^0+B_k 为这样的集合，它的每一个元恰为 x^0 和 B_k 中某元之和(x^0+B_k 也称为线性流形). 下一个定理刻划了共轭方向法的本质.

定理11.4(扩张子空间定理)　设 Q 为 n 阶对称正定阵，设 $d^0, d^1, \cdots, d^{n-1}$ 为一组非零的 Q 共轭向量. 任意取定 $x^0 \in R^n$，则按照

$$x^{k+1} = x^k + \alpha_k d^k, k = 0,1,\cdots,n-1 \tag{11.8}$$

其中

$$\alpha_k = -(d^k)^T g^k / (d^k)^T Q d^k \tag{11.9}$$

所产生的序列有这样的性质：x^k 使得

$$f(x) = \frac{1}{2} x^T Q x - b^T x \tag{11.10}$$

既在直线 $x = x^{k-1} + \alpha d^{k-1} (-\infty < \alpha < +\infty)$ 上取全局极小，又在线性流形 $x^0 + B_k$ 上取全局极小.

证　前一部分由 f 的构造和算法保证. 下证后一部分.

注意到 x^0+B_k 是 R^n 中的凸集且 $f(x)$ 是其上的凸函数，于是只要证，对$\forall \in x^0+B_k$ 都有

$$\nabla f(x^k)^T(x-x^k)=(g^k)^T(x-x^k)\geqslant 0$$

(可参见推论8.4)．再注意到对$\forall$ $x\in x^0+B_k$ 也有 $x-x^k\in B_k$，故只须证

$$(g^k)^Td^i=0, i=0,1,\cdots,k-1 \tag{11.11}$$

就足够了．用归纳法．

首先，由线搜索知，对 $i=0,1,\cdots,k-1$有

$$0=\frac{\mathrm{d}}{\mathrm{d}\alpha}f(x^i+\alpha d^i)\big|_{\alpha=\alpha_i}=\nabla f(x^i+\alpha_i d^i)^Td^i$$

也即

$$(g^{i+1})^Td^i=0, i=0,1,\cdots,k-1 \tag{11.12}$$

故(11.11)对 $k=1$是成立的．

设(11.11)成立．类(11.12)，先有

$$(g^{k+1})^Td^k=0 \tag{11.13}$$

又，由

$$x^{k+1}=x^k+\alpha_k d^k$$

结合(11.6)即推得

$$g^{k+1}=g^k+\alpha_k Qd^k \tag{11.14}$$

以 d^i，$i=0,1,\cdots,k-1$与之作内积得

$$(g^{k+1})^Td^i=(g^k)^Td^i+\alpha_k(d^i)^TQd^k, i<k \tag{11.15}$$

由归纳法之(11.11)成立知上式右之第一项为零，其第二项由共轭性知亦为零．故结合(11.13)和(11.15)得

$$(g^{k+1})^Td^i, i=0,1,\cdots,k$$

此即欲证．

由定理11.4和它的证明可得两个重要推论．

推论11.5 共轭方向法用于二次问题(11.1)所产生的 g^k 满足

$$(g^k)^Td^i=0, i=0,1,\cdots,k-1 \tag{11.16}$$

推论11.6(共轭方向法的二次终结性质) 共轭方向法用于相

应的二次问题(11.1),至多在 n 次得解.

证 若对某 $k<n$ 有 $g^k=\nabla f(x^k)=0$,则 x^k 为问题(11.1)的解. 否则,由推论11.5知 $(g^n)^T d^i=0, i=0,1,\cdots,n-1$. 但 $d^0,d^1,\cdots,d^{n-1}$ 线性无关,故 $g^n=\nabla f(x^n)=0$. 证毕.

对于已给的 n 阶对称正定阵,可以用不同的方法产生(不同的)Q 共轭向量,从而导致不同的共轭方向法. 由此,构造 Q 共轭方向的问题便显得重要了. 但构造对于二次问题具有良好性质的方法仅仅是研究问题的开始(二次问题总比较易于求解). 重要的是将所得的方法推广到用于一般非二次问题时也有某些较好的性质,例如比最速下降法有所改进,这是另一个重要的课题. 这两个问题归根结底仍然是寻找理想的搜索方向,这是本章余下部分的内容.

§11.2 共轭梯度法

与前面所述的共轭方向法不同,共轭梯度法不是一次便指定了 n 个 Q 共轭向量,而是在算法的迭代过程中逐次利用梯度的信息来构造共轭向量组. 因此,可以期望这种方法可以兼有共轭方法和梯度方法的优点.

共轭梯度法描述如下:任意取定初始点 $x^0\in R^n$,取 $d^0=-g^0=b-Qx^0$. 按下述公式进行迭代.

$$x^{k+1}=x^k+\alpha_k d^k \tag{11.17a}$$

其中

$$\alpha_k=-(d^k)^T g^k/(d^k)^T Q d^k \tag{11.17b}$$

以及

$$d^{k+1}=-g^{k+1}+\beta_k d^k \tag{11.18a}$$

其中

$$\beta_k=(g^{k+1})^T Q d^k/(d^k)^T Q d^k \tag{11.18b}$$

下面的定理给出了共轭梯度法用于二次问题的一些较好的性质,其中 $[g^0,g^1,\cdots,g^k]$ 表由 $\{g^0,g^1,\cdots,g^k\}$ 生成的子空间,等等.

定理11.7（共轭梯度定理） 共轭梯度法(11.17)、(11.18)是一种共轭方向法．如果在 x^k 处没有结束（即 $g^k=\nabla f(x^k)\neq 0$），那么

(i) $[g^0,g^1,\cdots,g^k]=[g^0,Qg^0,\cdots,Q^kg^0]$；

(ii) $[d^0,d^1,\cdots,d^k]=[g^0,Qg^0,\cdots,Q^kg^0]$；

(iii) $(d^k)^TQd^i=0,i<k$；

(iv) $\alpha_k=(g^k)^Tg^k/(d^k)^TQd^k$；

(v) $\beta_k=(g^{k+1})^Tg^{k+1}/(g^k)^Tg^k$．

证 如果在 x^k 处未结束，则 $g^i\neq 0,i\leqslant k,d^i\neq 0,i\leqslant k$．

我们先用归纳法同时证(i)～(iii)．对 $k=0$，显然它们都是真的．今假定(i)～(iii)对 $\leqslant k$ 成立．先由表达式 $g^i=Qx^i-b$ 以及由式(11.17a)得

$$g^{k+1}=g^k+\alpha_kQd^k \tag{11.19}$$

由归纳法假定，根据(i)和(ii)知 g^k、$Qd^k\in[g^0,Qg^0,\cdots,Q^{k+1}g^0]$．从而由(11.19)得

$$g^{k+1}\in[g^0,Qg^0,\cdots,Q^{k+1}g^0] \tag{11.20}$$

又由归纳法假定，根据(iii)知 $d^0,d^1,\cdots,d^k$ 为 Q 共轭向量．故可利用(11.16)（注意 $g^{k+1}\neq 0$）得

$$g^{k+1}\notin[d^0.d^1,\cdots,d^k]=[g^0,Qg^0,\cdots,Q^kg^0] \tag{11.21}$$

由(11.20)式和(11.21)式以及由(i)对 k 成立的假定，即推得(i)对 $k+1$ 成立．

由(ii)对 k 成立的假定，又已证(i)对 $k+1$ 成立，再利用(11.18a)即推得(ii)对 $k+1$ 亦成立．

为证(iii)对 $k+1$ 成立，以 Qd^i 对(11.18a)作内积得

$$(d^{k+1})^TQd^i=-(g^{k+1})^TQd^i+\beta_k(d^k)^TQd^i,i<k+1 \tag{11.22}$$

对 $i<k$，由归纳法假定知上式右端之第二项为零；同时

$$Qd^i\in[Qg^0,\cdots,Q^kg^0]\subseteq[d^0,d^1,\cdots,d^k],i<k \tag{11.23}$$

从而由(11.16)知 $g^{k+1}Qd^i=0,i<k$，总之，对 $i<k$ 成立着 $(d^{k+1})^TQd^i=0$．在 $i=k$ 时，由(11.18)直接计算得 $(d^{k+1})^TQd^k=0$．

这就证明了(iii).

将(11.18a)的上标减去1,然后与 g^k 作内积:

$$(g^k)^T d^k = (-g^k)^T g^k + \beta_{k-1}(g^k)^T d^{k-1} \tag{11.24}$$

上式右端之第二项为零,(iv)得证.

最后,由 $(g^{k+1})^T d^i=0, i<k+1$,以及由(i)、(ii)知 $g^k\in[d^0,d^1,\cdots,d^k]$,故 $(g^{k+1})^T g^k=0$. 用 g^{k+1} 与(11.19)作内积,得

$$\begin{aligned}(g^{k+1})^T Q d^k &= (g^{k+1})^T g^{k+1}/\alpha_k \\ &= (d^k)^T Q d^k \cdot (g^{k+1})^T g^{k+1}/(g^k)^T g^k\end{aligned}$$

后一等式是利用了(iv). 将其代入(11.18b)得(v). 证毕.

分析上定理可知,(iii)表明共轭梯度法也是一种共轭方向法,故它也具有二次终结性. (iv)和(v)表明(11.17a)和(11.18a)中的 α_k 和 β_k 的计算可以简化. 更重要的是 β_k 的新表达式为将共轭梯度法推广到用于非二次问题提供了方便(这点放在节末讲,注意 β_k 的新表达式中形式上已不含 Q!). 有趣的是,利用(ii)可以证明共轭梯度法用于二次问题在某种意义上为最优.

为此,设

$$f(x)=\frac{1}{2}x^T Q x - bx$$

的极小点为 x^*,利用(11.7)可得

$$f(x)=\frac{1}{2}(x-x^*)^T Q(x-x^*)-\frac{1}{2}(x^*)^T Q x^* \tag{11.25}$$

记

$$E(x)=\frac{1}{2}(x-x^*)^T Q(x-x^*) \tag{11.26}$$

则 $f(x)$ 和 $E(x)$ 有相同的极小点和相同的梯度.

设给定了初始点 x^0 和 k 个 Q 共轭方向. 则扩张子空间定理(定理11.4)告诉我们,由共轭梯度法得到的 x^{k+1} 使 $E(x)$ 在线性流形 x^0+B_{k+1} 上取最小. 定理11.7之(ii)告诉我们,x^{k+1} 有表达式

$$x^{k+1}=x^0+P_k(Q)g^0 \tag{11.27}$$

其中 P_k 为某个次数不超过 k 的多项式. 又, x^0+B_{k+1}中的任一元也可以由 $x^0+P_k(Q)g^0$表出;反之,任取一个次数不超过 k 的多项式 P_k, $x^0+P_k(Q)g^0$亦为 x^0+B_{k+1}中的元.

注意到 $g^0=Qx^0-b=Q(x^0-x^*)$,由(11.27)即得

$$\begin{aligned} x^{k+1}-x^* &= x^0+P_k(Q)Q(x^0-x^*)-x^* \\ &= [I+QP_k(Q)](x^0-x^*) \end{aligned} \tag{11.28}$$

将此式代入(11.26),得

$$E(x^{k+1})=\frac{1}{2}(x^0-x^*)^TQ[I+QP_k(Q)]^2(x^0-x^*) \tag{11.29}$$

将上面的分析综合起来,我们得到

定理11.8 共轭梯度法用于二次问题(11.1)产生的 x^{k+1}满足

$$E(x^{k+1})=\min_{P_k}\frac{1}{2}(x^0-x^*)^TQ[I+QP_k(Q)]^2(x^0-x^*) \tag{11.30}$$

其中极小是对所有次数不超过 k 的多项式 P_k 来取的.

设 Q 的特征值为 λ_i: $0<\lambda_1\leqslant\lambda_2\leqslant\cdots\leqslant\lambda_n$,其对应的规范化特征向量为 e_i, $i=1,2,\cdots,n$. 于是可设

$$x^0-x^*=\xi_1e_1+\xi_2e_2+\cdots+\xi_ne_n$$

如此就有

$$E(x^0)=\frac{1}{2}(x^0-x^*)^TQ(x^0-x^*)=\frac{1}{2}\sum_{i=1}^{n}\lambda_i(\xi_i)^2 \tag{11.31}$$

利用相同的方法可从(11.30)得到

$$\begin{aligned} E(x^{k+1}) &= \min_{P_k}\frac{1}{2}\sum_{i=1}^{n}[1+\lambda_iP_k(\lambda_i)]^2\lambda_i(\xi_i)^2 \\ &\leqslant \min_{P_k}\max_{\lambda_i}[1+\lambda_iP_k(\lambda_i)]^2\cdot\frac{1}{2}\sum_{i=1}^{n}\lambda_i(\xi_i)^2 \\ &= \min_{P_k}\max_{\lambda_i}[1+\lambda_iP_k(\lambda_i)]^2\cdot E(x^0) \end{aligned}$$

如此又有

定理11.9 在共轭梯度法中，我们有

$$E(x^{k+1}) \leqslant \min_{P_k} \max_{\lambda_i} [1 + \lambda_i P_k(\lambda_i)]^2 \cdot E(x^0) \quad (11.32)$$

其中极大是对 Q 的所有的特征值来取的；极小是对所有的次数不超过 k 的多项式来取的.

这个定理的一个简单的应用见习题3.

以上是对二次问题来讨论的. 将共轭梯度法略加修改可得下面的 FR(Fletcher-Reeves)方法. 它适用于一般非二次问题.

算法11.10(FR 方法)

设 $f: R^n \to R, f \in C^1$.

步0 任取初始点 $x^0 \in R^n$，指定一正整数 s，一般取 $s=n$. $k:=0$；

步1 令 $d^k = -\nabla f(x^k)$，若 $d^k=0$，停；否则，记

$$d^{k,0} = d^k, x^{k,0} = x^k, \text{对 } i = 0,1,\cdots,s-1$$

按如下格式进行迭代

(a) $x^{k,i+1} = x^{k,i} + \alpha_i d^{k,i}$；

$$f(x^{k,i+1}) = \min_{\alpha \geqslant 0} f(x^{k,i} + \alpha d^{k,i});$$

(b)若 $\nabla f(x^{k,i+1}) = 0$，停；否则

(c)除 $i=s-1$外，令

$$d^{k,i+1} = -\nabla f(x^{k,i+1}) + \beta_i d^{k,i}$$

其中

$$\beta_i = \nabla f(x^{k,i+1})^T \nabla f(x^{k,i+1}) / \nabla f(x^{k,i})^T \nabla f(x^{k,i});$$

步2 $x^{k+1} := x^{k,s}$，$k := k+1$，返步1.

容易看出，如果在步0中取 $s=1$，那么所得的方法就是最速下降法. FR 方法的收敛性可按最速下降法的收敛性的证明思路进行证明. 将 FR 方法适当修改，可得一个二次收敛的算法(条件与定理10.21的相同). 这里不介绍.

§11.3 拟 Newton 法的基本思想

我们已经指出，Newton 法的一种修改型为

$$d^k = -\nabla^2 f(x^k)^{-1}\nabla f(x^k) \tag{11.33}$$

$$x^{k+1} = x^k + \alpha_k d^k; f(x^{k+1}) = \min_{\alpha\geq 0} f(x^k + \alpha d^k) \tag{11.34}$$

在$\nabla f(x^k)\neq 0$时，为保证$f(x^{k+1})<f(x^k)$，希望$\nabla^2 f(x^k)$为正定. 但$\nabla^2 f(x^k)$有时是奇异的或接近奇异的. 又，即使$\nabla^2 f(x^k)$为正定，(11.33)中$\nabla^2 f(x^k)^{-1}$的计算也颇费时. 拟牛顿法就是为了克服这两个缺陷而设计的.

还是从正定二次问题着手，为阅读和书写方便，我们重新写出有关记号.

$$f(x) = \frac{1}{2}x^TQx - b^Tx \tag{11.35}$$

其中Q为n阶对称正定阵，记

$$g^k = \nabla f(x^k) = Qx^k - b \tag{11.36}$$

$$p^k = x^{k+1} - x^k = \alpha_k d^k \tag{11.37}$$

$$q^k = g^{k+1} - g^k = Q(x^{k+1} - x^k) = Qp^k \tag{11.38}$$

注意，求$f(x)$的极小等价于求解方程$Qx=b$，也可以说等价于求解Q^{-1}. 设已给定初始阵H_0，设法逐次对H_k作修正：

$$H_{k+1} = H_k + \Delta H_k \tag{11.39}$$

由(11.38)的启示，对任何k均有$p^k=Q^{-1}q^k$，故希望

$$H_{k+1}q^i = p^i, 0\leqslant i\leqslant k \tag{11.40}$$

如果依照某种迭代格式由(11.39)得到满足(11.40)的H_n，并且$p^i, i=0,1,\cdots,n-1$是线性无关的，则比较(11.38)和(11.40)即得$H_n=Q^{-1}$. 条件(11.40)称为拟 Newton 条件. 这条件在本章余下的两节中起重要作用.

利用p^k, q^k, H_k的信息来修正H_k，取

$$\Delta H_k = \Delta H(p^k, q^k, H_k)$$

使得$H_{k+1}=H_k+\Delta H_k$总满足(11.40)，利用不同的技巧来构造(不同的)ΔH_k，从而导致不同的方法，这些都是近四十多年来的工作. 注意，在拟 Newton 条件(11.40)中，我们并未考虑$p^i, q^i, i\leqslant k$是如何产生的. 但对于二次问题，或说对于由(11.36)～(11.38)所定义的p^i, q^i，它们满足

$$q^i = Qp^i, i = 0,1,2\cdots \tag{11.41}$$

例如，研究形如

$$H_{k+1} = H_k + r_k z^k (z^k)^T \tag{11.42}$$

的迭代的可能性. 其中 $r_k \in R, z^k \in R^n$.

在(11.40)中令 $i=k$ 得

$$p^k = H_{k+1} q^k = H_k q^k + r_k z^k (z^k)^T q^k \tag{11.43}$$

再用 q^k 与之作内积，得

$$(q^k)^T (p^k - H_k q^k) = r_k [(z^k)^T q^k]^2 \tag{11.44}$$

另一方面，利用(11.43)还可得

$$(p^k - H_k q^k)(p^k - H_k q^k)^T = r_k z^k (z^k)^T \cdot r_k [(z^k)^T q^k]^2 \tag{11.45}$$

由(11.45)解得 $r_k z^k (z^k)^T$，代入(11.42)，再利用(11.44)得

$$H_{k+1} = H_k + \frac{(p^k - H_k q^k)(p^k - H_k q^k)^T}{(q^k)^T (p^k - H_k q^k)} \tag{11.46}$$

当然要预设(11.46)中的分母不为零.

接下来，我们证明在(11.41)成立的条件下由(11.46)所定义的 H_{k+1} 满足(11.40). 用归纳法.

首先，由(11.46)能直接验证得

$$H_{k+1} q^k = p^k, \forall k$$

故(11.40)对 $k=0$ 成立.

假定对 H_k 有

$$H_k q^i = p^i, i \leqslant k-1$$

我们来证(11.40)成立. 对 $i<k$，由(11.46)，

$$H_{k+1} q^i = H_k q^i + y^k [(p^k)^T q^i - (q^k)^T H_k^T q^i]$$

其中 $y^k \in R^n$ 为某个向量. 若 $H_k^T = H_k$，即 H_k 为对称阵(这里要预先取定 H_0 为对称阵，则由(11.46)就能保证这一点)，则由归纳法假定有

$$H_{k+1} q^i = p^i + y^k [(p^k)^T q^i - (q^k)^T p^i], i < k \tag{11.47}$$

但是，利用(11.41)有

$$(p^k)^T q^i = (p^k)^T (Qp^i) = (Qp^k)^T p^i = (q^k)^T p^i$$

故由(11.47)得 $H_{k+1}q^i=p^i, i<k$. 由此结合上段之 $H_{k+1}q^k=p^k$ 知(11.40)成立.

迭代式(11.46)有两个严重的问题:分母为零怎么办?又,即使 H_k 为对称正定阵,H_{k+1}也可能非正定. 然而,(11.46)作为最早导出的这类公式有其特定的意义. 因为那里的修改阵 ΔH_k 的秩为1,故这方法也叫秩1法.

§11.4 DFP 法和 BFGS 法

本节所述的两种方法是目前广为流行的无约束最优化方法. 它们具有一些可人意的性质.

记 $H_{k+1}=H_k+\Delta H_k$. 仍从(11.40)出发,令 $i=k$ 时(11.40)成立,得

$$(H_k+\Delta H_k)q^k=p^k \tag{11.48}$$

等价地说,ΔH_k 应满足

$$\Delta H_k q^k=p^k-H_k q^k \tag{11.49}$$

1. DFP 法(Davidon-Fletcher-Powell)

依据(11.49),我们设想

$$\Delta H_k=p^k(u^k)^T-H_k q^k(v^k)^T \tag{11.50}$$

其中 $u^k, v^k\in R^n$ 为两个待定的向量. 如果能取得 u^k 和 v^k,使得

$$(u^k)^T q^k=(v^k)^T q^k=1 \tag{11.51}$$

则由(11.50)所确定的 ΔH_k 满足(11.49). 又,考虑到 ΔH_k 最好具有对称性,由(11.50)的启发,最简单的办法自然是令

$$u^k=\alpha_k p^k, v^k=\beta_k H_k q^k \tag{11.52}$$

其中 α_k 和 β_k 为待定的数. 为使由(11.52)式所确定的 u^k 和 v^k 满足(11.51),解得

$$\alpha_k=1/(p^k)^T q^k, \beta_k=1/(q^k)^T H_k q^k \tag{11.53}$$

依次利用(11.53)、(11.52)及(11.50),代入迭代式(11.39)得:

$$H_{k+1}=H_k+\frac{p^k(p^k)^T}{(p^k)^Tq^k}-\frac{H_kq^k(q^k)^TH_k}{(q^k)^TH_kq^k} \tag{11.54}$$

公式(11.54)称为 DFP 公式. 它首由 Davidon 提出原型，经 Fletcher 和 Powell 改进而成. 本节，我们恒设 H_0为对称阵.

下面，我们直接写出由 DFP 公式所得的算法，然后对算法进行讨论. 这算法适用于 R^n 中一般可微非线性函数的极小化问题.

算法11.11 (DFP 法)

步0　任给一初始对称正定阵 H_0(例如 $H_0=I$)，以及初始点 x^0，$k:=0$；

步1　若$\nabla f(x^k)=0$，停；否则，$d^k:=-H_k\nabla f(x^k)$；

步2　$x^{k+1}:=x^k+\alpha_kd^k$；$f(x^{k+1})=\min\limits_{\alpha\geqslant 0}f(x^k+\alpha d^k)$，

$p^k:=x^{k+1}-x^k$，

$q^k:=\nabla f(x^{k+1})-\nabla f(x^k)$；

步3　利用(11.54)计算 H_{k+1}，$k:=k+1$，返步1.

我们把算法的基本性质总结为如下的定理.

定理11.12　在算法11.11中，若$\nabla f(x^k)\neq 0$且 H_k 为正定，则

(i) $(q^k)^TH_kq^k>0$，$(p^k)^Tq^k>0$；

(ii) H_{k+1}为对称正定.

证　首先由步2之线搜索知

$$\nabla f(x^{k+1})^TH_k\nabla f(x^k)=0$$

(i) $$(q^k)^TH_kq^k=\nabla f(x^{k+1})^TH_k\nabla f(x^{k+1})+\nabla f(x^k)^TH_k\nabla f(x^k)>0$$

$$\begin{aligned}(p^k)^Tq^k&=\alpha_k(d^k)^Tq^k\\&=-\alpha_k\nabla f(x^k)^TH_k[\nabla f(x^{k+1})-\nabla f(x^k)]\\&=\alpha_k\nabla f(x^k)^TH_k\nabla f(x^k)>0(\text{注意 }\alpha_k>0)\end{aligned}$$

(ii) 任取 $x\in R^n$，$x\neq 0$，由(11.54)，

$$x^TH_{k+1}x=x^TH_kx+\frac{(x^Tp^k)^2}{(p^k)^Tq^k}-\frac{(x^TH_kq^k)^2}{(q^k)^TH_kq^k} \tag{11.55}$$

定义

$$a=H_k^{1/2}x,b=H_k^{1/2}q^k$$

则(11.55)可改写为

$$x^T H_{k+1} x = \frac{(a^T a)(b^T b) - (a^T b)^2}{b^T b} + \frac{(x^T p^k)^2}{(p^k)^T q^k} \quad (11.56)$$

上式右之第一项显然非负，第二项由$(p^k)^T q^k > 0$知亦非负. 若第一项为零，这意味着 $a=\beta b$ 对某个 $\beta \neq 0$成立(注意，$a \neq 0$)，由 a 及 b 之定义得 $x=\beta q^k$. 于是 $x^T p^k = \beta (q^k)^T p^k \neq 0$. 故总有 $x^T H_{k+1} x > 0$. 证毕.

在上面的证明中，$(p^k)^T q^k > 0$及$(q^k)^T H_k q^k > 0$是关键，而这一点恰由线搜索得以保证.

我们还要考虑 DFP 算法对于二次问题是否满足拟 Newton 条件. 事实上它还有更好的性质，这就是

定理11.13 如果 f 是具有正定 Hesse 阵 Q 的二次函数，那么，对于 DFP 法成立

$$(p^k)^T Q p^i = 0, i < k \quad (11.57)$$

$$H_{k+1} Q p^i = H_{k+1} q^i = p^i, i \leqslant k \quad (11.58)$$

证 在 $k=0$时，由 H_{k+1}的构造(11.54)知上两式为真. 下用归纳法证明我们的论断.

先证(11.57). 由(11.38)和(11.41)得

$$\begin{aligned} g^k &= g^{i+1} + (g^{i+2} - g^{i+1}) + \cdots + (g^k - g^{k-1}) \\ &= g^{i+1} + q^{i+1} + q^{i+2} + \cdots + q^{k-1} \\ &= g^{i+1} + Q(p^{i+1} + \cdots + p^{k-1}), i < k \end{aligned}$$

于是可得，对 $i<k$，

$$(p^i)^T g^k = (p^i)^T g^{i+1} + (p^i)^T Q(p^{i+1} + \cdots + p^{k-1}) = 0$$

上右端之第一项由线搜索知为零；第二项，将归纳法假定用于(11.57)，亦为零. 从而将归纳法假定用于(11.58)，得

$$0 = (p^i)^T g^k = [(p^i)^T Q H_k] g^k, i < k \quad (11.59)$$

于是，由 $p^k = -\alpha_k H_k g^k$ 及 $\alpha_k > 0$，由上式得$(p^i)^T Q p^k = 0, i<k$. 故(11.57)成立.

由归纳法假定，(11.58)对 k 成立，即 $H_k Q p^i = p^i, i \leqslant k-1$. 于是

$$(q^k)^T(H_kQp^i)=(q^k)^Tp^i=((p^k)^TQ)p^i=0, i\leqslant k-1 \quad (11.60)$$

先直接验算得 $H_{k+1}Qp^k=H_{k+1}q^k=p^k$. 再，对 $i\leqslant k-1$ 有

$$H_{k+1}Qp^i = H_kQp^i + \frac{p^k[(p^k)^TQp^i]}{(p^k)^Tq^k} - \frac{H_kq^k[(q^k)^TH_kQp^i]}{(q^k)^TH_kq^k}$$

由(11.60)知上两个方括号内消失，其右之第一项由归纳法知为 p^i ($i\leqslant k-1$). 故有 $H_{k+1}Qp^i=p^i, i\leqslant k$. 故(11.58)成立. 证毕.

(11.57)说明 DFP 法用于二次问题是一种共轭方向法，因此它具有二次终结性. 有趣的是，若在算法11.11中置 $H_0=I$，DFP 法对于二次问题就是共轭梯度法.

在适当的条件下，DFP 法用于非二次问题是全局收敛的，并且具有 n 步二次收敛速度. 即设 $\{x^k\}$ 为 DFP 法所产生的无穷点列，且 $x^k\to x^*$，则在适当的条件下，存在 $\eta>0$，使得

$$\|x_{k+n} - x^*\| \leqslant \eta\|x_k - x^*\|^2, k = 0,1,2,\cdots$$

这些证明都十分见长，故略.

2. BFGS 法(Broyden-Fletcher-Goldfarb-Shanno)

理论和实践表明，DFP 法是一种非常有效的无约束最优化方法，自从1959年发表以来，深受欢迎. 然而到了60年代后期，从计算实践中发现 DFP 法在数值稳定方面存在一定问题，从而人们又提出各种各样的修改算法. 其中被公认为具有更好数值稳定性的算法乃是 BFGS 法.

若在(11.50)中取

$$(u^k)^T = [1 + \beta(q^k)^TH_kq^k](p^k)^T/(p^k)^Tq^k - \beta(q^k)^TH_k \quad (11.61)$$

$$(v^k)^T = [1 - \beta(p^k)^Tq^k](q^k)^TH_k/(q^k)^TH_kq^k + \beta(p^k)^T \quad (11.62)$$

其中 β 是一个实参数. 直接验证可知(11.49)成立. 将上两式代入到(11.50)，再利用(11.39)得到公式

$$H_{k+1}= H_k + \frac{p^k(p^k)^T}{(p^k)^Tq^k} - \frac{H_kq^k(q^k)^TH_k}{(q^k)^TH_kq^k}$$
$$+ \beta\left\{\frac{(q^k)^TH_kq^k}{(p^k)^Tq^k}p^k(p^k)^T - p^k(q^k)^TH_k - H_kq^k(p^k)^T\right.$$

$$+\frac{(p^k)^T q^k}{(q^k)^T H_k q^k} H_k q^k (q^k)^T H_k \Big\} \tag{11.63}$$

若在(11.63)中取 $\beta=0$则得 DFP 法. 今取

$$\beta = 1/(p^k)^T q^k$$

代入(11.63),得

$$H_{k+1} = H_k + [\mu_k p^k (p^k)^T - p^k (q^k)^T H_k - H_k q^k (p^k)^T]/(p^k)^T q^k \tag{11.64a}$$

其中

$$\mu_k = 1 + \frac{(q^k)^T H_k q^k}{(p^k)^T q^k} \tag{11.64b}$$

(11.64)即 BFGS 公式. 将它代替算法11.11中步3之(11.54)即得 BFGS 算法. 它有与定理11.12和定理11.13完全一样的结论. 其证明亦相仿. 读者均可参见文献[1]、[2]或[23].

习　　题

1. 设 Q 是 n 阶对称正定阵,设 $y^0, y^1, \cdots, y^{n-1}$是 R^n 中线性无关的向量组. 证明,由 Gram-Schmidt 方法

$$d^0 = y^0,$$

$$d^{k+1} = y^{k+1} - \sum_{i=0}^{k} \frac{(y^{k+1})^T Q d^i}{(d^i)^T Q d^i} d^i$$

所递归定义的 $d^0, d^1, \cdots, d^{n-1}$构成一组非零的 Q 共轭向量.

2. 设 $f(x)=\frac{1}{2}x^TQx-b^Tx$,其中 Q 为 n 阶对称正定阵. 设 S_1和 S_2是 R^n 中两个不同的子空间,x^1和 x^2分别为 f 在 S_1和 S_2上的极小点. 证明,对$\forall\ d\in S_1\cap S_2$,$d$ 和 x^2-x^1是 Q 共轭的.

3. 在定理11.9的(11.32)中取 $k=0$,则对任意实数 α 都有

$$E(x^1) \leqslant \max_{\lambda\in[\lambda_1,\lambda_n]} (1+\alpha\lambda)^2 E(x^0) \tag{11.65}$$

其中 λ_1和 λ_n 分别表 Q 的最小和最大特征值. 上式实际上可解释为将最速下降法用于二次问题 $f(x)=\frac{1}{2}x^TQx-b^Tx$ 的一步的效果的估计. 由此,设$\{x^k\}$为由最速下降法用于二次问题所产生的点列,则也有

$$E(x^{k+1}) \leqslant \max_{\lambda\in[\lambda_1,\lambda_n]} (1+\alpha\lambda)^2 E(x^k) \tag{11.66}$$

试找出 α 的最好的值，使得(11.66)右端为最小.

4. 试证 FR 方法的收敛性.

5. 证明，若在 DFP 法中取 $H_0=I$，则 DFP 法用于二次问题就是共轭梯度法.

6. 在定理11.12之(ii)的证明中，所采用的 α_k 是使 $f(x^k+\alpha d^k)$ 精确地取得极小. 试说明凡使 $(p^k)^T q^k>0$ 的 α_k 都能确保 H_{k+1} 为正定. 说明这个结果的意义. 证明对于一个二次问题，任一个 $\alpha_k\neq0$ 都能产生正定的 H_{k+1}.

7. 设 A 为 $n\times n$ 阶非异阵，设 U 和 V 都是 $n\times m$ 阶阵 $(n\geqslant m)$. 证明，$(A+UV^T)$ 非异当且仅当 $(I+V^TA^{-1}U)$ 非异，并且

$$(A+UV^T)^{-1}=A^{-1}-A^{-1}U(I+V^TA^{-1}U)^{-1}V^TA^{-1} \tag{11.67}$$

8. 利用(11.67)推导 BFGS 公式(11.64)满足关系式

$$B_{k+1}=B_k+\frac{q^k(q^k)^T}{(p^k)^Tq^k}-\frac{B_kp^k(p^k)^TB_k}{(p^k)^TB_kp^k} \tag{11.68}$$

其中 $B_k=H_k^{-1}$，$B_{k+1}=H_{k+1}^{-1}$. 试将(11.68)与 DFP 公式(11.54)作比较.

第十二章　线性逼近法

本章和下章，我们考虑约束最优化，所论的问题是

$$\min_{x\in S} f(x) \tag{12.0}$$

其中 S 是 R^n 中的子集，f 是 R^n 中或 S 中的实值函数.

设已给初始点 $x^0\in S$，依照某种迭代格式从 $x^k\in S$ 寻找一个 $x^{k+1}\in S$，使得 $f(x^{k+1})<f(x^k)$. 按这种方法构得序列 $\{x^k\}$，从而可望求解所论的问题. 本章以及下章的部分都是这种思想的具体实现.

§12.1　可行方向法

设 $f(x)$ 在 R^n 上是连续可微的. 对于(12.0)，设已有了初始点 $x^0\in S$. 我们记得，满足

$$\nabla f(x^0)^T d < 0 \tag{12.1}$$

的 d^0 便是 f 在 x^0 处的下降方向. 如果这个 d^0 还是可行的，即存在 $\bar{\alpha}_0>0$ 使得

$$x^0 + \alpha d^0 \in S, \forall\ \alpha \in [0,\bar{\alpha}_0] \tag{12.2}$$

则称 d^0 为 f 在 x^0 处的一个可行下降方向.

可行方向法分二步：先找一个 x^0 处的可行下降方向 d^0，再从 x^0 出发沿 d^0 对 f 在可行的条件下求极小 x^1，或找 α_0：

$$x^1 = x^0 + \alpha_0 d^0, f(x^1) = \min_{0\leqslant\alpha\leqslant\bar{\alpha}_0} f(x^0 + \alpha d^0)$$

Zoudendijk 对于线性约束下的非线性规划

$$\text{(LNP)}\qquad \begin{aligned} &\min f(x) \\ &\text{s.t. } a_i^T x + b_i = 0, i=1,2,\cdots,m \\ &\qquad a_j^T x + b_j \leqslant 0, j=m+1,\cdots,p \end{aligned}$$

其中 $a_i, a_j \in R^n, b_i, b_j \in R$，提出了一个寻找可行下降方向的策略. 设已给(LNP)的可行点 x^0. 由于约束全是线性的，容易验证，x^0处的可行方向集为

$$Z(x^0) = \{d \in R^n : a_i^T d = 0, i = 1,2,\cdots,m; a_j^T d \leqslant 0, j \in J(x^0)\}$$

其中

$$J(x^0) = \{j : a_j^T x^0 + b_j = 0, j = m+1,\cdots,p\}$$

那么可取 d^0为线性规划

$$\min_{d, d \in Z(x^0), |d_l| \leqslant 1, l=1,\cdots,n} \nabla f(x^0)^T d \tag{12.3}$$

的解，其中 d_l 表 d 的第 l 个分量，$|d_l| \leqslant 1, l=1,\cdots,n$ 是附加的约束，以保证所涉的线性规划有解.

如果由(12.3)所得的 d^0满足(12.1)，则它必是 f 在 x^0处的一个可行下降方向. 于是从点 x^0沿 d^0可得一个可行点 x^1使 $f(x^1) < f(x^0)$. 如果所得的 d^0不满足(12.1)，即 $\nabla f(x^0)^T d^0 = 0$，依(12.3)的定义得

$$\{d : \nabla f(x^0)^T d < 0\} \cap Z(x^0) = \varnothing$$

利用定理8.16知此时 x^0已为(LNP)的 $K-T$ 点. 这样，我们已经描述了文献中最早出现的可行方向法的完整的一步.

为了确保算法的收敛性，Frank，Wolfe 在寻找可行下降方向时，不同于上面的(12.3)，代之以求解(LP_{x^0})：

$$(LP_x) \qquad \begin{aligned} &\min_d \nabla f(x)^T d \\ &\text{s.t.}\ a_i^T d = 0, i=1,\cdots,m \\ &\qquad a_j^T(x+d) + b_j \leqslant 0, j=m+1,\cdots,p \\ &\qquad |d_l| \leqslant M, l=1,\cdots,n \end{aligned}$$

其中 $M>0$预先给定. 注意，(LP_x)是将所有的约束一起考虑进去. 并且，(12.3)的最优值为零，当且仅当(LP_{x^0})的最优值为零.

现在，我们叙述方法.

算法12.1(Frank-Wolfe 方法)

步0　对(LNP)，给定 $M>0$，任给可行点 x^0，$k := 0$；

步1　求解线性规划(LP_{x^k})，得解 d^k；

步2　若$\nabla f(x^k)^T d^k=0$，停；否则

步3　$x^{k+1}=x^k+\alpha_k d^k$：$f(x^{k+1})=\min\limits_{0\leqslant\alpha\leqslant 1} f(x^k+\alpha d^k)$，$k:=k+1$，返步1.

现在给出 Frank-Wolfe 方法的收敛性.

定理12.2　设 $f\in C^1$，则算法12.1有如下性质：

(i) 若算法终止到某 k_0，则 x^{k_0} 为(LNP)的 K-T 点；

(ii) 若算法产生一个无穷点列$\{x^k\}$且设$\{x^k\}$有界，则$\{x^k\}$的任一聚点都是(LNP)的 K-T 点；

(iii) 若 $f(x)$为凸函数，则上述(i)、(ii)所得之点为(LNP)的全局极小点.

证　(i)由前面的分析得到，(iii)由(i)或(ii)及(LNP)为凸规划得到. 我们用全局收敛性定理证(ii). 定义

$\Gamma(x)=\{x:x$ 为(LNP) 的 K-T 点$\}$

$\qquad=\{x:x$ 为(LNP) 的可行点且(LP_x) 的为优值为零$\}$

映射 $A=SG$，其中 S 即步3，是受限制的线搜索，它是闭的(见定理10.20类型 I)；$G:R^n\to R^{2n}$由(LP_x)定义(见步1)，它也是闭的(见下引理). 于是由定理10.8知 $A=SG$ 是闭的. 再定义关于 A 和 Γ 的下降函数 $\omega(x)=f(x)$，则利用全局收敛性定理知(ii)为真. 证毕.

引理12.3　设 $f\in C^1$，则由(LP_x)定义的 $G(x)=(x,d)$是闭的.

证　设 d 由

$$(\mathrm{LP}_x)'\qquad \min_d \nabla f(x)^T d$$
$$\text{s.t.}\ \ a_j^T(x+d)+b_j\leqslant 0,\ j=1,\cdots,r$$

所定义. 为证得 $G(x)=(x,d)$为闭，我们只须证由$(\mathrm{LP}_x)'$所定义的 $\hat{G}(x)=(x,d)$为闭则可. 事实上，将 $a_i^T d=0$化为等价的 $a_i^T d\leqslant 0$，$-a_i^T d\leqslant 0$；同时将 $|d_l|\leqslant M$ 化为等价的 $d_l\leqslant M$，$-d_l\leqslant M$ 后，(LP_x)即成了$(\mathrm{LP}_x)'$.

设 $x^k\to x$，$d^k\to d$，d^k 是$(\mathrm{LP}_{x^k})'$的解. 欲证 $\hat{G}$ 为闭就是证 d 也是$(\mathrm{LP}_x)'$的解. 记

$$J(x^k)=\{j:a_j^T(x^k+d^k)+b_j=0,j=1,\cdots,r\}$$

d^k 是$(LP_{x^k})'$的解的充要条件是，存在 $\mu_j^k\geqslant0,j\in J(x^k)$，使得

$$\nabla f(x^k)+\sum_{j\in J(x^k)}\mu_j^k a_j=0 \tag{12.4}$$

再记

$$J(x)=\{j:a_j^T(x+d)+b_j=0,j=1,\cdots,r\}$$

由 $x^k+d^k\to x+d$ 知，对充分大的 k，恒有 $J(x^k)\subseteq J(x)$. 故(12.4)可改写为

$$\nabla f(x^k)+\sum_{j\in J(x)}\mu_j^k a_j=0 \tag{12.5}$$

其中对 $j\in J(x)\setminus J(x^k)$，补设 $\mu_j^k=0$. 于是仍有 $\mu_j^k\geqslant0,j\in J(x)$. 利用代数的处理，对(12.5)还可设：$\mu_j^k\geqslant0,j\in J(x)$，且相应于 $\mu_j^k>0$ 的全部 a_j 是线性无关的(可参见定理1.10中(i)的证明).

由于 $J(x)$是有限集，故存在 $J(x)$的一个子集 $\hat{J}(x)$，使得在序列$\{\mu_j^k:\mu_j^k\geqslant0,j\in J(x)\}$中有一个子列

$$\{\mu_j^{k_i}:\mu_j^{k_i}>0,j\in\hat{J}(x);\mu_j^{k_i}=0,j\in J(x)\setminus\hat{J}(x)\}$$

如此，由(12.5)又得

$$\nabla f(x^{k_i})+\sum_{j\in\hat{J}(x)}\mu_j^{k_i}a_j=0 \tag{12.6}$$

且由上段末的说明可设$\{a_j:j\in\hat{J}(x)\}$是线性无关的.

由$\{a_j:j\in\hat{J}(x)\}$的线性无关性及条件 $x^k\to x$，从(12.6)得 $\mu_j^{k_i}\to\mu_j^*\geqslant0,j\in\hat{J}(x)$，且

$$\nabla f(x)+\sum_{j\in\hat{J}(x)}\mu_j^*a_j=0 \tag{12.7}$$

也即(对 $j\in J(x)\setminus\hat{J}(x)$，补设 $\mu_j^*=0$)

$$\nabla f(x)+\sum_{j\in J(x)}\mu_j^*a_j=0 \tag{12.8}$$

(12.8)就是 d 是$(LP_x)'$的解的充要条件. 引理证毕.

应该指出，在算法12.1中，如果求解的问题是无约束问题，即(LNP)中的约束全消失，那么对应于(LP_x)中的约束只剩下$|d_l|\leqslant M,l=1,\cdots,n$. 所得的算法与最速下降法相类似(具有带限制的线搜索). 于是我们只能期望本算法仅具有线性收敛速度.

§ 12.2 线性化方法

我们试图将算法12.1的步1和步3加以合并. 如果步1中求得的 d^k 恰满足

$$f(x^k + d^k) < f(x^k) \tag{12.9}$$

由于 $x^k + d^k$ 对(LNP)是可行的，则令

$$x^{k+1} = x^k + d^k$$

或许可得一个稍为简单些的方法.

注意到若 $f \in C^1$，则对小的 d^k 有

$$f(x^k + d^k) \approx f(x^k) + \nabla f(x^k)^T d^k \tag{12.10}$$

如果 $\nabla f(x^k)^T d^k < 0$，则(12.9)对于小的 d^k 会成立. 于是若在算法12.1中将步1之(LP_{x^k})中之 $|d_l| \leqslant M$ 改为 $|d_l| \leqslant \delta_k, l = 1, \cdots, n$，并取 δ_k 足够小则能保证(12.9)成立. 然而这样的改动势必会使新的算法收敛过慢. 于是想到当 $f(x^k)$ 与 $f(x^k + d^k)$ 有较大差别时，应将 δ_k 增大. 这些想法可用下面的办法来实现：预先给定 ρ_1, ρ_2，：$0 < \rho_1 < \rho_2 < 1$，考虑

$$f(x^k + d^k) - f(x^k) \leqslant \rho_1 \nabla f(x^k)^T d^k \tag{12.11}$$

$$f(x^k + d^k) - f(x^k) \leqslant \rho_2 \nabla f(x^k)^T d^k \tag{12.12}$$

如果对于 $|d_l| \leqslant \sigma_k$，解对应的(LP_{x^k})所得之 d^k 能满足(12.12)，则置 $x^{k+1} = x^k + d^k$，并且此时说明 f 在 x^k 处沿方向 d^k 有较大下降的趋势. 下一步将 σ_k(按某种法则)增大到 σ_{k+1}，类似地求解对应的($LP_{x^{k+1}}$)；又若这 d^k 不满足(12.11)，这说明 $f(x^k + d^k)$ 与 $f(x^k)$ 比较未下降或下降得过少，则可将 σ_k(按某种法则)缩小后，重解对应的(LP_{x^k})；对剩下的情况，则令 $\sigma_{k+1} = \sigma_k, x^{k+1} = x^k + d^k$，继续迭代. 总之，一个合意的 $x^{k+1} = x^k + d^k$ 是通过(在必要时)不断地调整 σ_k，使(12.11)总成立.

由此构造的算法主要是通过求 $f(x)$ 的一次近似式，即(12.10)之右端的极小来完成的，故可称之为线性化方法. 从另一个角度来看，这个方法的手段是通过不断地调整 σ_k 来实现的，使

得有关线性规划的约束区域能保证得到的 d^k 使 f 有足够的下降，故又被称为信赖域法(Trust Region Algorithm)，英文缩写为TRA.

为阅读和书写方便，我们写出

(LNP) $$\min f(x)$$
$$\text{s.t. } a_i^T x+b_i=0, i=1,\cdots,m$$
$$a_j^T x+b_j\leqslant 0, j=m+1,\cdots,p$$

以及一个新的辅助线性规划

($LP_{x,\sigma}$) $$\min_d \nabla f(x)^T d$$
$$\text{s.t. } a_i^T d=0, i=1,\cdots,m$$
$$a_j^T(x+d)+b_j\leqslant 0, j=m+1,\cdots,p$$
$$|d_l|\leqslant\sigma, l=1,\cdots,n$$

算法12.4(线性化方法)

步0 预先指定 ρ_1,ρ_2,c_1,c_2: $0<\rho_1<\rho_2<1, c_2\geqslant 2, 0<c_1<1$，以及一个较大的 $M>0$，任取一个可行的 x^0，$\sigma_0:=M$，$k:=0$；

步1 求解(LP_{x^k,σ_k})，得解 d^k；

步2 若 $\nabla f(x^k)^T d^k=0$，停；否则

步3 考虑

$$f(x^k+d^k)-f(x^k)\leqslant\rho\nabla f(x^k)^T d^k \tag{12.13}$$

(a)若(12.13)对 $\rho=\rho_1$ 不成立，则 $\sigma_k:=c_1\sigma_k$，重解(LP_{x^k,σ_k})得解仍记为 d^k，用新的 d^k 代替原有的 d^k 重新执行步3；

(b)若(12.13)对 $\rho=\rho_1$ 成立但对 $\rho=\rho_2$ 不成立，则 $\sigma_{k+1}:=\sigma_k$，转步4；

(c)若(12.13)对 $\rho=\rho_2$ 成立，则 $\sigma_{k+1}:=\min\{c_2\sigma_k,M\}$，转步4；

步4 $x^{k+1}:=x^k+d^k$，$k:=k+1$. 返步1.

下面考虑收敛性.

定理12.5 设 $f\in C^1$，则由算法12.4求解(LNP)，有

(i) 若算法终止于某 k_0，则 x^{k_0} 即为(LNP)的K-T点；

(ii) 若算法产生一个无穷序列 $\{x^k\}$，则它的任一聚点都是

(LNP)的 K-T 点；

(iii) 若 $f(x)$为凸函数，则由(i)及(ii)所得之点都是(LNP)的全局极小点.

证 (i)、(iii)的证明同定理12.2，下证(ii).

我们只在 $x^k \to \bar{x}$，即$\{x^k\}$本身收敛的情况下进行证明. 首先，由$\{x^k\}$有界及 $f \in C^1$ 知，对$\forall\ \varepsilon > 0$，存在 $\delta_1 > 0$，使得只要$\|d^k\| = \|d^k\|_\infty < \delta_1$，就有

$$f(x^k + d^k) - f(x^k) \leqslant \nabla f(x^k)^T d^k + \varepsilon \|d^k\| \quad (12.14)$$

其次，若 $\bar{x}$ 非(LNP)的 K-T 点. 记 $\bar{d}^1$为($LP_{\bar{x},1}$)的解. 则由 $\bar{x}$ 非(LNP)的 K-T 点知$-\nabla f(\bar{x})^T \bar{d}^1 \triangleq 2L > 0$；再记 $d^{k,1}$为($LP_{x^k,1}$)的解，再由 $f \in C^1$ 知，存在 $\delta_2 > 0$，只要$\|x^k - \bar{x}\| < \delta_2$，就有$-\nabla f(x^k)^T d^{k,1} \geqslant L$(参见引理12.3)，于是

$$-\nabla f(x^k)^T d^k \geqslant -\nabla f(x^k)^T (\|d^k\| \cdot d^{k,1}) \geqslant L \cdot \|d^k\| \quad (12.15)$$

结合(12.14)和(12.15)可知，只要$\|d^k\| < \delta_1$，$\|x^k - \bar{x}\| < \delta_2$，就有

$$\frac{f(x^k + d^k) - f(x^k)}{-\nabla f(x^k)^T d^k} \leqslant \frac{\nabla f(x^k)^T d^k + \varepsilon \|d^k\|}{-\nabla f(x^k)^T d^k} \leqslant -1 + \varepsilon / L$$

在上式中取 $\varepsilon = L(1 - \rho_2)$，注意到

$$x^k \to \bar{x}, \text{从而 } d^k \to 0 \triangleq \bar{d} \quad (12.16)$$

故存在 K_1，只要 $k \geqslant K_1$就有

$$f(x^k + d^k) - f(x^k) \leqslant \rho_2 \nabla f(x^k)^T d^k \quad (12.17)$$

最后，(12.17)表明算法之第3步在 $k \geqslant K_1$后总执行(c)，故又存在 $K_2 \geqslant K_1$，只要 $k \geqslant K_2$就有 $\sigma_k = M$. 于是在 $k \geqslant K_2$时，算法12.4步1之(LP_{x^k,σ_k})就是算法12.1中的(LP_{x^k}). 利用引理12.3及(12.16)知 $\bar{d} = 0$就是($LP_{\bar{x},M}$)的解；但由$\nabla f(\bar{x})^T \bar{d} = 0$知 $\bar{x}$ 是(LNP)的 K-T 点. 此与 $\bar{x}$ 非(LNP)的 K-T 点的假设矛盾. 证毕.

请读者检查以上证明的每一步，然后想一想，如果只取$\{x^k\}$的一个聚点来证会发生什么困难?这个问题困扰着人们近二十年. 我们要在下一节对一个更为一般的问题作正面的解答.

§12.3 似线性化方法

我们可以将上节的思路加以延伸. 为此, 令

$$f=(f_1,f_2,\cdots,f_r)^T$$

其中 $f_i\in C^1, i=1,\cdots,r$ 为 R^n 上的实值函数. 令 φ 为 R^r 上的凸函数, 令 S 为 R^n 中的闭凸集. 考虑问题

(GNP) $$\min_{x\in S}\varphi(f(x))$$

注意, 若取 S 如上节的线性约束集, 取 $r=1$, 取 $\varphi=t: R\to R$, 则 (GNP) 就是上节的 (LNP).

今将 f 线性化, 记

$$G(x;d)=\varphi(f(x)+f'(x)d)$$
$$\triangleq\varphi(f_1(x)+\nabla f_1(x)^Td,\cdots,f_r(x)+\nabla f_r(x)^Td) \tag{12.18}$$

并构造如下的辅助规划

($\mathrm{GLP}_{x,\sigma}$) $$\min_{d:\|d\|\leqslant\sigma,x+d\in S}G(x;d)$$

这规划中, $x\in S$ 为固定, $d\in R^n$ 是变量. 现在我们可以描述求解 (GNP) 的算法.

算法12.6(似线性化方法)

步0 预先指定 ρ_1,ρ_2,c_1,c_2: $0<\rho_1<\rho_2<1, 0<c_1<1, c_2\geqslant2$, 以及某个较大的 $M>0$. 任取 $x^0\in S, \sigma_0:=M, k:=0$;

步1 求解 ($\mathrm{GLP}_{x^k,\sigma_k}$), 得解 d^k;

步2 若 $G(x^k;0)-G(x^k;d^k)=0$, 停; 否则

步3 考虑

$$\varphi(f(x^k+d^k))-\varphi(f(x^k))\leqslant\rho(G(x^k;d^k)-G(x^k;0)) \tag{12.19}$$

(a) 若 (12.19) 对 $\rho=\rho_1$ 不成立, 则 $\sigma_k:=c_1\sigma_k$, 重解 ($\mathrm{GLP}_{x^k,\sigma_k}$) 得解仍记为 d^k, 用新的 d^k 代替原有的 d^k 重新执行步3;

(b) 若 (12.19) 对 $\rho=\rho_1$ 成立但对 $\rho=\rho_2$ 不成立, 则 $\sigma_{k+1}:=\sigma_k$,

转步4；

(c)若(12.19)对 $\rho=\rho_2$ 成立，则 $\sigma_{k+1}:=\min\{c_2\sigma_k, M\}$，转步4；

步4 $x^{k+1}:=x^k+d^k, k:=k+1$，返步1.

以下对算法的性质进行讨论.

定义12.7 设 $x\in S$，说 x 是(GNP)的(广义)K-T 点，若存在某个 $\sigma>0$ 使得 $d=0$ 为($\mathrm{GLP}_{x,\sigma}$)的解. (GNP)的 K-T 点的全体记为 Γ.

引理12.8 设 $f\in C^1$，φ 为 R^r 中的凸函数，则 $G(x;d)$ 关于 d 是凸函数.

证 任意 $d^1, d^2\in R^n$，$\forall\ \lambda\in[0,1]$，

$$\begin{aligned}
G(x;\lambda d^1+(1-\lambda)d^2) &= \varphi(f(x)+f'(x)(\lambda d^1+(1-\lambda)d^2)) \\
&= \varphi[\lambda(f(x)+f'(x)d^1)+(1-\lambda)(f(x)+f'(x)d^2)] \\
&\leqslant \lambda\varphi(f(x)+f'(x)d^1)+(1-\lambda)\varphi(f(x)+f'(x)d^2) \\
&= \lambda G(x;d^1)+(1-\lambda)G(x;d^2)
\end{aligned}$$

注 这个引理的一个直接应用就是算法12.6中的辅助规划($\mathrm{GLP}_{x^k,\sigma_k}$)总是凸规划，故其极小是全局性的.

以下，记($\mathrm{GLP}_{x,\sigma}$)的解为 $d_{x,\sigma}$.

引理12.9 设 $f\in C^1$，φ 为 R^r 上的凸函数且 S 为 R^n 中之凸集. 若 $\bar{x}\in S$ 但 $\bar{x}\notin\Gamma$，则存在 $L>0$ 及 $\delta_1>0$，只要 $\|x-\bar{x}\|\leqslant\delta_1$，就有

$$G(x;0)-G(x;d_{x,1})\geqslant L \tag{12.20}$$

证 设 $\bar{x}\in S$ 但 $\bar{x}\notin\Gamma$，由定义12.7知

$$G(\bar{x};0)-G(\bar{x};d_{\bar{x},1})\triangleq 2L>0 \tag{12.21}$$

取 $\bar{\lambda}=1/(1+\|d_{\bar{x},1}\|)$，记

$$\tilde{x}=\bar{x}+d_{\bar{x},1}(\in S) \tag{12.22}$$

由 S 为凸，就有

$$x+\bar{\lambda}(\tilde{x}-x)\in S, \forall\ x\in S \tag{12.23}$$

再设 $x\in S:\|x-\bar{x}\|\leqslant 1$，于是

$$\|\bar{\lambda}(\tilde{x}-x)\|=\left\|\frac{1}{1+\|d_{\bar{x},1}\|}(\bar{x}+d_{\bar{x},1}-x)\right\|$$

$$\leqslant \frac{1}{1+\|d_{\bar{x},1}\|}(\|\bar{x}-x\|+\|d_{\bar{x},1}\|)$$
$$\leqslant 1 \tag{12.24}$$

由(12.23)、(12.24)及 $d_{x,1}$的定义知，对$\forall\ x\in S$：$\|x-\bar{x}\|\leqslant 1$，有

$$G(x;\bar{\lambda}(\tilde{x}-x))-G(x;d_{x,1})\geqslant 0 \tag{12.25}$$

利用此式以及 $G(x;d)$关于 d 为凸，再利用(12.22)，有

$$\begin{aligned}
&G(x;0)-G(x;d_{x,1})\\
&=[G(x;0)-G(x;\bar{\lambda}(\tilde{x}-x))]+[G(x;\bar{\lambda}(\tilde{x}-x))-G(x;d_{x,1})]\\
&\geqslant[G(x;0)-G(x;\bar{\lambda}(\tilde{x}-x))]\\
&\geqslant\bar{\lambda}[G(x;0)-G(x;(\tilde{x}-x))]\\
&=\bar{\lambda}\{[G(x;0)-G(\bar{x};0)]+[G(\bar{x};0)-G(\bar{x};d_{\bar{x},1})]\\
&\qquad+[G(\bar{x};\tilde{x}-\bar{x})-G(x;\tilde{x}-x)]\}
\end{aligned}$$

由 f'的连续性及φ的凸性(从而连续)知，上式之花括号中之第一、三项均可充分小，只要$\|x-\bar{x}\|$充分小. 对其第二项用(12.21)得(12.20). 证毕.

由引理12.9可直接得到一个有用的推论.

推论12.10 设 $f\in C^1$，φ为 R^r 中之凸函数且 S 为 R^n 中之闭凸集，则 Γ 为闭集.

证 利用上引理及定义12.7可证 Γ(相对于 S)的补为开集.

引理12.11 设 $f\in C^1$，φ为 R^r 中之凸函数，则$\forall\ x\in S$，$\forall\ \sigma>0$，有

$$G(x;0)-G(x;d_{x,\sigma})\geqslant\min\{1,\sigma\}(G(x;0)-G(x;d_{x,1})) \tag{12.26}$$

证 对 $\sigma\geqslant 1$上式显然成立. 若$0<\sigma<1$，则

$$\begin{aligned}
G(x;0)-G(x;d_{x,\sigma})&\geqslant G(x;0)-G(x;\sigma d_{x,1})\\
&\geqslant\sigma(G(x;0)-G(x;d_{x,1}))
\end{aligned}$$

上之第一个不等式直接由定义得到；第二个不等式由 $G(x;d)$关于 d 的凸性得到，故(12.26)亦成立. 证毕.

由引理12.9和引理12.11直接得如下结论.

引理12.12 设 $f\in C^1$，φ为 R^r 中之凸函数且 S 为 R^n 中之凸

集. 若 $\bar{x}\in S$ 但 $\bar{x}\notin\Gamma$, 则存在 $\delta_1>0$ 及 $L>0$, 只要由算法给出的 x^k 满足 $\|x^k-\bar{x}\|\leqslant\delta_1$, 就有

$$G(x^k;0)-G(x^k;d^k)\geqslant L\cdot\min\{1,\sigma_k\} \qquad (12.27)$$

引理12.13 设 $f\in C^1$, φ 为 R^r 中的凸函数, 则对 $\forall\ x\in S$, 有

$$\varphi(f(x+d))=G(x;d)+o(\|d\|) \qquad (12.28)$$

证 固定 x. 由 φ 为凸, 故 φ 满足局部 Lipschitz 条件. 记 L_f 为 $f(x)$ 的某个领域内 Lipschitz 常数, 则对充分小的 d, 有

$$\begin{aligned}|\varphi(f(x+d))-G(x;d)|&=|\varphi(f(x+d))-\varphi(f(x)+f'(x)d)|\\&\leqslant L_f\cdot\|f(x+d)-f(x)-f'(x)d\| \qquad (12.29)\end{aligned}$$

由此式及条件 $f\in C^1$ 即推得(12.28). 证毕.

引理12.14 设 $f\in C^1$, φ 为 R^r 中之凸函数且 S 为 R^n 中之凸集. 若 $\bar{x}\in S$ 但 $\bar{x}\notin\Gamma$, 则存在 $\delta_1>0$, $\delta_2>0$, 只要 $\|x^k-\bar{x}\|\leqslant\delta_1$, $\|d^k\|\leqslant\delta_2$, 就有

$$\varphi(f(x^k+d^k))-\varphi(f(x^k))\leqslant\rho_2(G(x^k;d^k)-G(x^k;0)) \qquad (12.30)$$

证 由引理12.12之(12.27)知, 若 $\|x^k-\bar{x}\|\leqslant\delta_1$ 及 $\|d^k\|\leqslant1$, 则

$$G(x^k;0)-G(x^k;d^k)\geqslant L\|d^k\| \qquad (12.31)$$

由此及引理12.13, 并注意到 $\varphi(f(x^k))=G(x^k;0)$, 得

$$\begin{aligned}\frac{\varphi(f(x^k))-\varphi(f(x^k+d^k))}{G(x^k;0)-G(x^k;d^k)}&=\frac{G(x^k;0)-G(x^k;d^k)+o(\|d^k\|)}{G(x^k;0)-G(x^k;d^k)}\\&=1+\frac{o(\|d^k\|)}{\|d^k\|} \qquad (12.32)\end{aligned}$$

由 $\rho_2<1$, 取 $\delta_2>0$ 适当小, 即由(12.32)得(12.30). 证毕.

由于在得到(12.30)的证明过程中并不涉及算法的具体迭代, 故对 $\bar{x}\in S$ 但 $\bar{x}\notin\Gamma$, 在(12.30)中以 $\bar{x}$ 代替 x^k, 以 $d_{\bar{x},\sigma}$ 代替 d^k, 则对 $\forall\ \sigma$: $0\leqslant\|d_{\bar{x},\sigma}\|\leqslant\delta_2$, 新(12.30)依然成立. 于是, 我们得到

推论12.15 设 $f\in C^1$, φ 为 R^r 中之凸函数且 S 为 R^n 中之闭凸集, 则算法12.6是可行的.

证 在算法步3之(a)中, σ_k 会不断地缩小. 此时, 由 $x^k\in S$ 但 $x^k\notin\Gamma$, 视 k 为固定, 视 x^k 为上引理中之 $\bar{x}$, 于是经过有限次的 σ_k

的缩小后，(12.30)必成立. 故算法必转入(b)或(c)，证毕.

Ortega，Rheinboldt 定义的伪凸函数如下，在凸集 $X\subseteq R^n$ 上的实值函数 g 称为伪凸的，如果对$\forall\ x^1,x^2\in X$ 并且满足 $g(x^1)>g(x^2)$，都存在 α 和 τ：$\alpha>0$，$0<\tau\leqslant 1$（一般 α 和 τ 都依赖于 x^1 和 x^2），使得对每个$0\leqslant\theta\leqslant\tau$，满足

$$g(x^1+\theta(x^2-x^1))\leqslant g(x^1)-\theta\alpha \tag{12.33}$$

我们称 g 为在"Ortega-Rheinboldt"意义下的伪凸.

注意，若 g 为凸，则(12.33)必成立.

定理12.16 设 $f\in C^1$，φ 为 R^r 上的凸函数且 S 为 R^n 中之凸集，则

(i) (GNP)的极小点都是它的 K-T 点；

(ii) 若 $\varphi(f(x))$在 S 上为在 Ortega-Rheinboldt 意义下伪凸，则(GNP)的 K-T 点都是它的全局极小点.

证 (i) 若 $\bar{x}\in S$ 但 $\bar{x}\notin\Gamma$，则在(12.30)中令 x^k 为 $\bar{x}$，取 σ：$0<\sigma\leqslant\delta_2$，$\delta_2$如引理12.14，可得新(12.30). 由引理12.12之(12.27)知新(12.30)右端为负，这表明 $\bar{x}$ 非(GNP)之极小点.

(ii) 任取定 $\bar{x},x\in S$，$\bar{x}\in\Gamma$. 若 $\varphi(f(\bar{x}))>\varphi(f(x))$，则依伪凸条件，在(12.33)中分别视 x^1,x^2为 $\bar{x},x$，则存在 $\alpha>0$，使得对$\forall\ \theta\in(0,\tau]$，均成立

$$\begin{aligned}-\alpha&\geqslant\frac{\varphi(f(\bar{x}+\theta(x-\bar{x})))-\varphi f(f(\bar{x}))}{\theta}\\&=\frac{G(\bar{x};\theta(x-\bar{x}))-G(\bar{x};0)}{\theta}+\frac{o(\theta)}{\theta}\end{aligned} \tag{12.34}$$

上式中的等式由(12.28)式得到. 又，由 $\bar{x}\in\Gamma$ 知上式最右端之第一项恒非负. 令 $\theta\to0^+$可知(12.34)式不成立. 这个矛盾说明了$\varphi(f(\bar{x}))\leqslant\varphi(f(x))$. 证毕.

注 这个定理说明了 K-T 点的定义(定义12.7)的合理性.

引理12.17 设 $f\in C^1$，φ 为 R^r 中之凸函数且 S 为 R^n 中之闭凸集. 设算法12.6产生一个无穷点列$\{x^k\}$. 若$\{x^k\}$有一聚点 $\bar{x}\notin\Gamma$，则必存在一个 $\hat{\delta}>0$：可从$\{x^k\}$中选出一个收敛子列$\{x^{k_j}\}$，$x^{k_j}\to\hat{x}$，

$\hat{x}\in S$，且使得 $\sigma_{k_j}\geqslant\hat{\delta}$ 对一切 j 成立，并且还有 $\hat{x}\notin\Gamma$.

证 在引理12.14中先将 $\delta_1>0$适当缩小，使得凡满足$\|x-\bar{x}\|\leqslant\delta_1$的 x 都有 $x\notin\Gamma$(由推论12.10以及 $\bar{x}\notin\Gamma$ 知这是办得到的)；再将 $\delta_2>0$适当缩小，使得

$$2\delta_2\leqslant\delta_1 \text{ 且 } c_2\delta_2<M$$

于是，对$\|x^k-\bar{x}\|\leqslant2\delta_2$，$\|d^k\|\leqslant\delta_2$依然有(12.30).

今设$\{x^{k_t}\}$为$\{x^k\}$的收敛子列，且 $x^{k_t}\to\bar{x}$. 由 S 为闭知 $\bar{x}\in S$. 不妨设对一切 t 有$\|x^{k_t}-\bar{x}\|\leqslant\delta_2$. 从$\{x^{k_t}\}$中任取一元 x^{k_T}：

(i) 若 $\sigma_{k_T}\geqslant\delta_2$，则选 x^{k_T}；

(ii) 若 $\sigma_{k_T}<\delta_2$，注意到$\|d^{k_T}\|\leqslant\sigma_{k_T}$，以及

$$\|x^{k_T}-\bar{x}\|\leqslant\delta_2 \tag{12.35}$$

故有

$$\begin{aligned}\|x^{k_T+1}-\bar{x}\|&=\|x^{k_T}+d^{k_T}-\bar{x}\|\\&\leqslant\|x^{k_T}-\bar{x}\|+\|d^{k_T}\|<2\delta_2\end{aligned}$$

据此可设

$$\text{对 } s=0,1,\cdots,\hat{s}\text{，有 }\delta_{k_T+s}<\delta_2\text{ 但 }\sigma_{k_T+\hat{s}+1}\geqslant\delta_2 \tag{12.36a}$$

以及

$$\text{对 } s=1,2,\cdots,\bar{s}\text{，有}\|x^{k_T+s}-\bar{x}\|<2\delta_2\text{但}\|x^{k_T+\bar{s}+1}-\bar{x}\|\geqslant2\delta_2 \tag{12.36b}$$

再记

$$s^*=\min\{\hat{s},\bar{s}\}$$

则对 $s=0,1,\cdots,s^*$ 均有(注意到还有(12.35))

$$\|d^{k_T+s}\|\leqslant\sigma_{k_T+s}<\delta_2,\ \|x^{k_T+s}-\bar{x}\|<2\delta_2$$

由第一段的叙述，可重写(12.30)为

$$\begin{aligned}&\varphi(f(x^{k_T+s}+d^{k_T+s}))-\varphi(f(x^{k_T+s}))\\&\leqslant\rho_2(G(x^{k_T+s};d^{k_T+s})-G(x^{k_T+s};0))\end{aligned}$$

于是此时算法步3执行(c)，即

$$\text{对 } s=0,1,\cdots,s^*\text{，有 }\sigma_{k_T+s+1}:=c_2\sigma_{k_T+s} \tag{12.37}$$

今证 $\hat{s}\leqslant\bar{s}$. 若 $\hat{s}>\bar{s}$，则上式中有 $s^*=\bar{s}$. 由于 $c_2\geqslant2$，上式表明以

$\|x-x^{k_T+\bar{s}+1}\|\leqslant\sigma_{k_T+\bar{s}+1}$所示的邻域应包含 x^{k_T+s}，$s=0,1,\cdots,\bar{s}$. 于是结合(12.35)有

$$\begin{aligned}\sigma_{k_T+\bar{s}+1} &\geqslant \|x^{k_T}-x^{k_T+\bar{s}+1}\| \\ &\geqslant \|x^{k_T+\bar{s}+1}-\bar{x}\|-\|x^{k_T}-\bar{x}\| \\ &\geqslant 2\delta_2-\delta_2=\delta_2\end{aligned}$$

另一方面，由(12.36)知 $\hat{s}>\bar{s}$ 又意味着 $\sigma_{k_T+\bar{s}+1}<\delta_2$. 这个矛盾便证明了 $\hat{s}\leqslant\bar{s}$.

接下来证明 $\hat{s}$ 为有限. 若 $\hat{s}$ 为无限，则由 $\bar{s}\geqslant\hat{s}$ 知(12.37)中有 $s^*=+\infty$，即得 $\sigma_{k_T+s}=c_2^s\sigma_{k_T}$对所有 $s=0,1,2,\cdots$成立. 这又表明 $\sigma_{k_T+s}\geqslant\delta_2$对适当大的 s 总成立. 这与 $\hat{s}$ 为无限相矛盾.

既然 $\hat{s}$ 为有限且 $\hat{s}\leqslant\bar{s}$. 此时选 $x^{k_T+\hat{s}}$. 由(12.36)和(12.37)即得

$$\|x^{k_T+\hat{s}}-\bar{x}\|\leqslant 2\delta_2\leqslant\delta_1 \tag{12.38}$$

$$\sigma_{k_T+\hat{s}}=\sigma_{k_T+\hat{s}+1}/c_2\geqslant\delta_2/c_2 \tag{12.39}$$

注意，情况(i)中所选的 x^{k_T}也满足上二式.

综上(i)和(ii)知，逐一考察$\{x^{k_i}\}$中的每一元，则可从$\{x^k\}$中选得一子列$\{x^{k_j}\}$，并且 x^{k_j}和 σ_{k_j}满足对应的(12.38)和(12.39). 视$\{x^{k_j}\}$本身收敛，记 $\hat{\delta}=\delta_2/c_2$，结合本证明之第一段的论述知结论为真. 证毕.

定理12.18(似线性化方法的收敛性) 设 $f\in C^1$，φ为 R^r 中的凸函数且 S 为 R^n 中之闭凸集，则算法12.6是可行的；并且

(i) 若算法终止于某 x^k，则 $x^k\in\Gamma$；

(ii) 若算法产生一个无穷点列$\{x^k\}$，则对于它的任一聚点 $\bar{x}$ 都有 $\bar{x}\in\Gamma$；

(iii) 若 $\varphi(f)$在 S 上为在 Ortega-Rheinboldt 意义下伪凸，则(i)中之 x^k 或(ii)中之 $\bar{x}$ 都是(GNP)的全局极小点.

证 可行性即推论12.15. (i)由算法得到保证，(iii)由(i)、(ii)及定理12.16之(ii)得到，下证(ii).

若 $\bar{x}\notin\Gamma$，则由引理12.17知，可取得$\{x^k\}$中的收敛子列$\{x^{k_j}\}$：$x^{k_j}\rightarrow\hat{x}$，$\hat{x}\in S$，$\hat{x}\notin\Gamma$；并且存在 $\hat{\delta}>0$使得 $\sigma_{k_j}\geqslant\hat{\delta}$ 对一切 j 成立. 由

引理12.12知存在 $\delta_1>0$及 $L>0$,只要$\|x^{k_j}-\hat{x}\|\leqslant\delta_1$就有

$$
\begin{aligned}
&G(x^{k_j};0)-G(x^{k_j};d^{k_j})\\
&\geqslant L\cdot\min\{1,\sigma_{k_j}\}\geqslant L\cdot\min\{1,\hat{\delta}\}\\
&\triangleq\delta_0>0
\end{aligned}
\tag{12.40}
$$

另一方面,由算法结合(12.40)有

$$
\begin{aligned}
&\varphi(f(x^{k_j}))-\varphi(f(x^{k_j+1}))\\
&\geqslant\rho_1\cdot(G(x^{k_j};0)-G(x^{k_j};d^{k_j}))\geqslant\rho_1\cdot\delta_0>0
\end{aligned}
\tag{12.41}
$$

由$\{\varphi(f(x^k))\}$的单调下降性及 $x^{k_j}\to\hat{x}$ 知上式之最左端的极限为零,这又表明(12.41)对充分大的 j 不成立.定理证毕.

注 综观以上的证明知,若在算法12.6步3之(a)中

$$\text{将 }\sigma_k:=c_1\sigma_k\text{改为 }\sigma_k:=c_1\cdot\|d^k\| \tag{12.42}$$

同时(或者)在步3之(b)中

$$\text{将 }\sigma_{k+1}:=\sigma_k\text{改为 }\sigma_{k+1}:=\|d^k\| \tag{12.43}$$

以上一切结论都成立.

又,若定义 $\varphi_i:R^r\to R,i=1,2,3$,为

$$\varphi_1(t)=\max_{1\leqslant j\leqslant r}\{t_j\}$$

$$\varphi_2(t)=\sum_{j=1}^{r}|t_j|$$

$$\varphi_3(t)=\sum_{j=1}^{r}(t_j)^2$$

它们都是 R^r 中的凸函数.对应于(GNP)则有

$$\min_{x\in S}\max_{1\leqslant j\leqslant r}\{f_j(x)\} \tag{12.44}$$

$$\min_{x\in S}\sum_{j=1}^{r}|f_j(x)| \tag{12.45}$$

$$\min_{x\in S}\sum_{j=1}^{r}[f_j(x)]^2 \tag{12.46}$$

它们在文献中分别称为极大的极小问题,(非线性)l_1问题,以及最小平方和问题.在 $S=R^n$ 或 S 由线性约束刻划时,由算法12.6求解这些问题都比较方便.

对于带有非线性约束的问题，可用本章及下章的方法结合求解.

习 题

1. 利用算法12.1求解

$$\min\ 4(x_1)^2+(x_2-2)^2$$
$$\text{s.t.}\ \ |x_1|\leqslant 2,\ |x_2|\leqslant 1$$

取 $x^0=(-2,-1)^T$，$M=4$. 并作示意图.

2. 利用算法12.4求解上题中的问题. 取 x^0，M 同上题，然后取 $c_1=1/2$，$c_2=2$，$\rho_1=1/4$，$\rho_2=2/3$. 并作示意图.

3. 利用算法12.1求解

$$\min\ f(x)=\frac{4}{3}[(x_1)^2-x_1x_2+(x_2)^2]^{3/4}-x_3$$
$$\text{s.t.}\ \ x_1,x_2,x_3\geqslant 0$$

取 $x^0=(0,1/4,1/2)^T$，$M=5$.

4. 利用(12.3)表示的规划代替算法12.1步1中的规划后所得的算法，就是 Zoudendijk 最初提出的可行方向法. 试用这个方法求解上题的规划. 取 x^0 和 M 与上题同. 看看会发生什么问题.

5. 设 $\phi(x,y)$. $R^n\times R^n\to R$，$S\subseteq R^n$. 对 $\forall\ x\in S$，考虑

$$(\mathrm{P}_x)\qquad \inf_{y:x+y\in S,\|y\|\leqslant 1}\phi(x,y)$$

记它的最优值为 $v(x)$. 证明

(i) 若 S 为凸集且 ϕ 为上半连续，则 v 为上半连续；

(ii) 若 S 为凸集且 ϕ 为连续，则 v 亦为连续；

(iii) 若 S 为闭集且 ϕ 为下半连续，则 v 为下半连续.

注　这里的结论是引理12.3和引理12.9的推广.

第十三章　罚函数法

求解约束优化问题的另一个重要途径是化约束问题为无约束问题，这就是著名的罚函数法．罚函数法也叫 SUMT 方法，即序列无约束极小化方法（Seqential Unconstrained Minimization Techniques）.

§13.1　外部罚函数法

1. 方法

仍考虑问题

(P)
$$\min_{x\in S} f(x)$$

其中 f 为 R^n 上的连续函数，$S\subseteq R^n$．假定还存在 R^n 上的续函数 $P(x)$，其满足

$$P(x)\begin{cases}=0, & 若\ x\in S\\ >0, & 若\ x\notin S\end{cases} \tag{13.1}$$

借此定义一个 R^n 中含参变量 μ 的函数

$$q(\mu;x)=f(x)+\mu P(x) \tag{13.2}$$

并称之为问题(P)的增广目标函数．由(13.1)和(13.2)可知

$$q(+\infty;x)=\begin{cases}f(x), & 若\ x\in S\\ +\infty, & 若\ x\notin S\end{cases}$$

于是，形式地可将(P)化为无约束问题

$$\min_{x\in S} f(x)=\min_{x\in R^n} q(+\infty;x) \tag{13.3}$$

实际上，(13.3)右端无法求解，于是想到用一列单调上升趋于正无穷的 μ_k 代之，即求解(序列的)无约束问题

$$\min_{x\in R^n} q(\mu_k,x) \tag{13.4}$$

希望求得的一系列 x^k 能提供(P)的解的信息.

算法13.1(外部罚函数法)

对规划(P),设已给由(13.1)所定义的 $P(x)$.

步0　取定一严格单调上升的正数列 $\mu_k, k=0,1,2,\cdots$,且 $\mu_k\to+\infty$, $k:=0$;

步1　求解(13.4),得解 x^k;

步2　若 $x^k\in S$,停,x^k 为(P)之解;否则

步3　$k:=k+1$,返步1.

本章所述之极小皆指全局极小. 首先,我们有

引理13.2　若存在一个 μ_{k0},使得 $\min\limits_{x\in R^n} q(\mu_{k0};x)$ 的解 $x^{k_0}\in S$,则 x^{k_0} 为(P)之解.

证　任取 $x\in S$,由 $P(x)$之定义知

$$\begin{aligned}f(x) &\geqslant \min_{x\in S} f(x) = \min_{x\in S}\{f(x)+\mu_{k0}P(x)\}\\ &\geqslant \min_{x\in R^n}\{f(x)+\mu_{k0}P(x)\} = f(x^{k_0})\end{aligned}$$

证毕.

如先前一样,如果算法得到一个无穷点列,则问题总要复杂一些.

引理13.3　设算法13.1给出了一个无穷点列$\{x^k\}$,则下列关系式总成立:

(i) $q(\mu_k;x^k)\leqslant q(\mu_{k+1};x^{k+1})$;　(13.5)

(ii) $P(x^k)\geqslant P(x^{k+1})$;　(13.6)

(iii) $f(x^k)\leqslant f(x^{k+1})$.　(13.7)

证　(i)

$$\begin{aligned}q(\mu_{k+1};x^{k+1}) &= f(x^{k+1})+\mu_{k+1}P(x^{k+1})\\ &\geqslant f(x^{k+1})+\mu_kP(x^{k+1})\\ &\geqslant \min_{x\in R^n}\{f(x)+\mu_kP(x)\}\\ &= q(\mu_k;x^k)\end{aligned}$$

(ii) 由定义得

$$f(x^{k+1})+\mu_{k+1}P(x^{k+1})\leqslant f(x^k)+\mu_{k+1}P(x^k)\tag{13.8}$$

$$f(x^k)+\mu_kP(x^k)\leqslant f(x^{k+1})+\mu_kP(x^{k+1})\tag{13.9}$$

两式相加后,整理得

$$(\mu_{k+1}-\mu_k)P(x^{k+1}) \leqslant (\mu_{k+1}-\mu_k)P(x^k)$$

由 $\mu_{k+1}-\mu_k>0$，由上式即推得(13.6).

(iii) 由(13.9)及(ii)即得. 证毕.

定理13.4(外部罚函数法的收敛性)　用算法13.1求解(P)有如下性质：

(i)若算法终止于某 k_0，则 x^{k_0} 为(P)之解；

(ii)若算法产生一个无穷点列 $\{x^k\}$，则 $\{x^k\}$ 的任一聚点都是(P)之解.

证　(i)即引理13.2.

(ii)设 $\{x^{k_i}\}\subseteq\{x^k\}$，$x^{k_i}\to\bar{x}$. 任取定 $x\in S$，则有

$$\begin{aligned} f(x) &= f(x)+\mu_k P(x) \\ &\geqslant f(x^k)+\mu_k P(x^k) \geqslant f(x^k) \end{aligned} \tag{13.10}$$

故有

$$f(x) \geqslant f(\bar{x}), \forall\, x\in S \tag{13.11}$$

同时，由(13.10)还知

$$f(x^{k_i})+\mu_{k_i}P(x^{k_i}) \leqslant f(x) \tag{13.12}$$

这意味着 $P(\bar{x})=0$. 不然的话，(13.12)对充分大的 μ_{k_i} 将不成立. 由 $P(\bar{x})=0$ 得 $\bar{x}\in S$. 由此结合(13.11)知 $\bar{x}$ 为(P)之解. 证毕.

注意，由上面的证明知可在(13.10)中取 $x=\bar{x}$，从而得 $f(\bar{x})\geqslant f(x^k)$，故一般地说有 $x^k\notin S$. 并且 $\{x^k\}$ 是从 S 的外部去逼近(P)的解的.

推论13.5　设算法13.1产生一个无穷点列 $\{x^k\}$. 若 $\{x^k\}$ 有一聚点 x^*，则 x^* 为(P)之解. 且

(i) $\lim\limits_{k\to\infty} f(x^k)=f(x^*)$；

(ii) $\lim\limits_{k\to\infty} q(\mu_k;x^k)=f(x^*)$；

(iii) $\lim\limits_{k\to\infty} \mu_k P(x^k)=0$.

证明留给读者. 注意，(iii) 表明在 $k\to\infty$ 时，惩罚项的作用逐渐消失.

下面考虑 $P(x)$ 的构造.

情况 I　$S_1=\{x\in R^n; h_i(x)=0, i=1,\cdots,m\}$.

取

$$P_1(x)=\sum_{i=1}^{m}|h_i(x)|^{\alpha}, \alpha>0$$

若 $\alpha>0$ 且 h_i 为连续，则 $P_1(x)$ 连续；若取 $\alpha>1$ 且 h_i 可微，则 $P_1(x)$ 亦可微.

情况 II　$S_2=\{x\in R^n; g_j(x)\leqslant 0, j=1,\cdots,p\}$.

取

$$P_2(x)=\sum_{j=1}^{p}[\max\{0,g_j(x)\}]^{\beta}, \beta>0$$

对于 β 的讨论与上面的相同.

情况 III　$S=S_1\cap S_2$.

取

$$P(x)=P_1(x)+P_2(x)$$

例13.6　$\min f(x)=(x)^2$, s.t. $1-x\leqslant 0, x\in R$.

显然，$x^*=1, f(x^*)=1$. 取

$$P(x)=[\max\{0,1-x\}]^2,$$

则

$$q(\mu;x)=f+\mu P=\begin{cases}(x)^2, & 若\ x\geqslant 1\\ (x)^2+\mu(1-x)^2, & 若\ x<1\end{cases}$$

在 R 中极小化 $q(\mu;x)$：

$\forall\ \mu$：$\min\limits_{x\geqslant 1} q(\mu;x)=1$，极小点 $\bar{x}=1$

$\forall\ \mu>0$：$\min\limits_{x<1} q(\mu;x)=\min\limits_{x<1}((x)^2+\mu(1-x)^2)$，故极小点为 $\hat{x}=\mu/(1+\mu)<1$. 同时

$$q(\mu,\hat{x})=\left(\frac{\mu}{1+\mu}\right)^2+\mu\left(\frac{\mu}{1+\mu}-1\right)^2=\frac{\mu}{1+\mu}<1$$

综合得

$$q(\mu_k;x^k)=\min_{x\in R} q(\mu_k;x)=\mu_k/(1+\mu_k)$$

$$x^k=\mu_k/(1+\mu_k)$$

于是

$$q(\mu_k;x^k)\to 1=f(x^*);x^k\to 1=x^*$$

可从图13.1中验证以上有关理论：

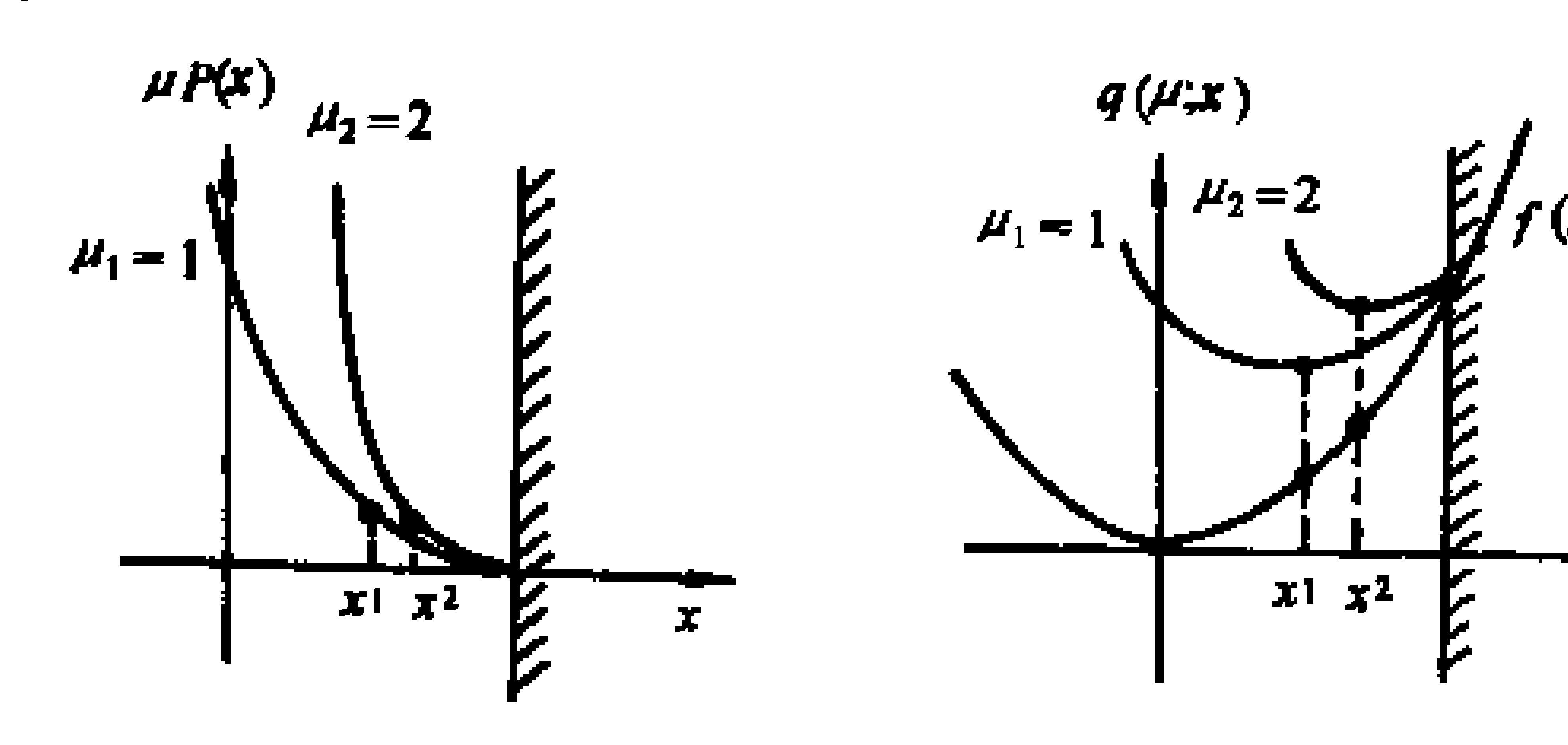


图 13.1

2. 外部罚函数的性质

本节，我们考虑外部罚函数的性质. 为简便计，仅考虑等式约束的情况

(EP) $\qquad \min f(x)$

$\qquad\qquad$ s. t. $h_i(x)=0, i=1,\cdots,m$

假定(EP)有解 x^*，$f, h_i\in C^2$ 且 x^* 是约束集的正则点，即 $\nabla h_i(x^*), i=1,\cdots,m$，线性无关.

取

$$q(\mu_k;x)=f(x)+\mu_k\sum_{i=1}^{m}|h_i(x)|^2 \tag{13.13}$$

记 $h=(h_1,\cdots,h_m)^T$，则上式可写成紧凑的形式：

$$q(\mu_k;x)=f(x)+\mu_k h(x)^T h(x).$$

(i)设 x^k 为$\min\limits_{x\in R^n} q(\mu_k;x)$的解，则必有

$$\begin{aligned}0&=\nabla_x q(\mu_k;x)|_{x=x^k}\\&=\nabla f(x^k)+2\mu_k\nabla h(x^k)h(x^k)\end{aligned} \tag{13.14}$$

其中$\nabla h(x^k)$的定义见§8.2. 若定义

$$\lambda^k=2\mu_k h(x^k)(\in R^m) \tag{13.15}$$

则(13.14)可改写为

$$\nabla f(x^k)+\nabla h(x^k)\lambda^k=0 \tag{13.16}$$

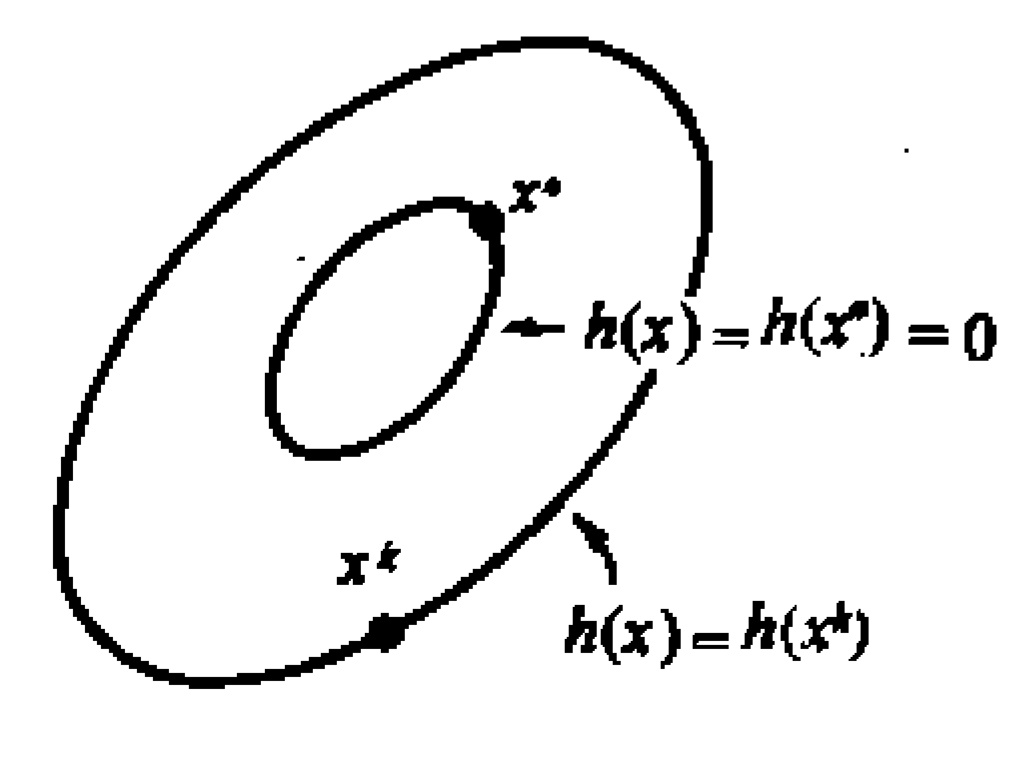


图 13.2

(13.16)恰表明 x^k 是问题

$$(\text{EP}_k)\qquad \min f(x)$$
$$\text{s.t.}\ h(x)=h(x^k)$$

的 K-T 点. 几何上的解释见图13.2.

(ii) 由于 $\nabla h(x^*)$ 满秩，故当 x^k 充分接近于 x^* 时，$\nabla h(x^k)$ 亦满秩，从而由(13.16)解得

$$\lambda^k = -[\nabla h(x^k)^T \nabla h(x^k)]^{-1} \nabla h(x^k)^T \nabla f(x^k) \quad (13.17)$$

令 $k\to\infty$，得 $\lambda^k\to\lambda^*$，其中

$$\lambda^* = -[\nabla h(x^*)^T \nabla h(x^*)]^{-1} \nabla h(x^*)^T \nabla f(x^*)$$

设 $\lambda_i^k, \lambda_i^*, i=1,\cdots,m$ 分别为 λ^k 和 λ^* 的第 i 个分量. 又若 $\lambda_i^*\neq 0$，则由(13.15)，$\lambda_i^k=2\mu_k h_i(x^k)\to\lambda_i^*\neq 0$. 这又意味着 $h_i(x^k)$ 趋于零的速度可表为 $1/\mu_k$.

(iii) 现在考虑 Hesse 阵

$$\nabla_{xx}^2 q(\mu_k; x^k) = \nabla^2 f(x^k) + 2\mu_k \nabla h(x^k) \nabla h(x^k)^T + 2\mu_k \nabla^2 h(x^k) h(x^k)$$

或记为

$$\nabla_{xx}^2 q(\mu_k; x^k) = L(x^k) + 2\mu_k \nabla h(x^k) \nabla h(x^k)^T \quad (13.18)$$

其中

$$L(x^k) = \nabla^2 f(x^k) + 2\mu_k \nabla^2 h(x^k) h(x^k)$$

注意到 $\lambda^k = 2\mu_k h(x^k)\to\lambda^*$，则

$$L(x^k) \to \nabla^2 f(x^*) + \nabla^2 h(x^*)\lambda^* = L(x^*) \quad (13.19)$$

由于

$$\nabla h(x^k)\nabla h(x^k)^T \to \nabla h(x^*)\nabla h(x^*)^T$$

以及由正则性知 $\nabla h(x^*)\nabla h(x^*)^T$ 有 m 个特征值大于零，有 $n-m$ 个特征值为零. 令再假定(EP)在 x^* 处满足二阶充分条件，即由(13.19)所定义的 $L(x^*)$ 在

$$M(x^*) = \{d\in R^n; \nabla h(x^*)^T d = 0\}$$

上为正定，那么由(13.18)可以推得当 $\mu_k\to\infty$ 时，$\nabla_{xx}^2 q(\mu_k; x^k)$ 有 m 个特征值趋于正无穷，而另外的 $n-m$ 个特征值为有限的正数，也

就是说，μ_k 越大，Hesse 阵越接近于病态.

求解无约束优化问题的比较理想的方法为第十一章所述的一些拟 Newton 法. 而这类方法的效率强烈地依赖于 Hesse 阵的性态. 外部罚函数法的脆弱之处即在于此. 尽管如此，对于带有非线性约束的规划问题的求解，外部罚函数法仍不失为一个有效的方法而广被采用.

§ 13.2 内部罚函数法

1. 方法

对于问题

(P) $$\min_{x\in S} f(x)$$

言，其中 f 是 S 上的连续函数，自然也会想到是否能从约束集的内部去逼近(P)的解. 为使方法可行以及考虑到收敛性，我们对集合 S 作如下的规定：

(i) S 有一个非空的内部，即 $\text{int } S\neq\varnothing$；

(ii) $\text{cl int } S=S$.

注意，(ii)说明了 S 是闭的，且 S 中的点可由其内部的点来逼近.

今在 $\text{int } S$ 上定义一个连续函数 $B(x)$，它满足

(i) $B(x)\geqslant 0, \forall\, x\in \text{int } S$；

(ii) $x\rightarrow S$ 的边界时，恒有 $B(x)\rightarrow+\infty$.

定义增广目标函数(其中 $\mu>0$ 为参数)

$$r(\mu;x)=f(x)+\frac{1}{\mu}B(x) \tag{13.20}$$

取 $0<\mu_1<\mu_2<\cdots$，且 $\mu_k\rightarrow\infty$，求解

$$\min_{x\in \text{int}S} r(\mu_k;x) \tag{13.21}$$

记解为 x^k. 注意，(13.21)形式上虽是约束问题，但因为 $r(\mu_k;x)$ 对任意的 $\mu_k>0$ 在 S 的边界的附近可以任意大，从而产生屏障，故在实际计算时其相当于一个无约束问题. 在迭代开始时应先求得 S

的一个内点，即 $x^0\in \text{int } S$；然后从 x^0 出发求解 $\min\limits_{x\in \text{int } S} r(\mu_1;x)$，得到 $x^1\in \text{int } S$，如此等等.

内部罚函数有如下性质，我们把证明留给读者.

引理13.7 设求解(13.21)得 $\{x^k\}$，则

(i) $r(\mu_k;x^k)\geqslant r(\mu_{k+1};x^{k+1})$； (13.22)

(ii) $B(x^k)\leqslant B(x^{k+1})$； (13.23)

(iii) $f(x^k)\geqslant f(x^{k+1})$. (13.24)

定理13.8(内部罚函数法的收敛性) 设求解(13.21)得一个无穷序列 $\{x^k\}$，则它的任一聚点都是(P)的解.

证 不妨设 $\{x^k\}$ 本身收敛，且 $x^k\to\bar{x}$，由于 $x^k\in \text{int } S$ 及 S 为闭，故 $\bar{x}\in S$.

今任取定 $x\in \text{int } S$，则对 $\forall\ k$，有

$$r(\mu_k;x^k)=f(x^k)+\frac{1}{\mu_k}B(x^k)$$

$$\leqslant f(x)+\frac{1}{\mu_k}B(x) \tag{13.25}$$

由 $\mu_k>0, B(x^k)\geqslant 0$，由上式即得

$$f(x^k)\leqslant f(x)+\frac{1}{\mu_k}B(x)$$

从而就有

$$f(\bar{x})\leqslant f(x)+\frac{1}{\mu_k}B(x) \tag{13.26}$$

在上式中令 $k\to\infty$，得

$$f(\bar{x})\leqslant f(x), \forall\ x\in \text{int } S \tag{13.27}$$

但 S 中的任一点都可由 int S 中的点去逼近，以及 f 在 S 上是连续的，故由(13.27)即推得

$$f(\bar{x})\leqslant f(x), \forall\ x\in S \tag{13.28}$$

由 $\bar{x}\in S$，上试表明 $\bar{x}$ 为(P)之解. 证毕.

下一个结论亦留给读者证明.

推论13.9 设规划(13.21)产生了一个无穷序列 $\{x^k\}$. 若 x^* 为它的一个聚点，则 x^* 为(P)的解；且

(i) $\lim_{k\to\infty} f(x^k)=f(x^*)$;

(ii) $\lim_{k\to\infty} r(\mu_k;x^k)=f(x^*)$;

(iii) $\lim_{k\to\infty}\frac{1}{\mu_k}B(x^k)=0$.

内部罚函数法可用于求解只含有不等式约束的问题：

(IP) $$\min_{x\in S} f(x)$$

其中

$$S=\{x\in R^n : g_j(x)\leqslant 0, j=1,\cdots,p\}$$

假定 g_j 连续，且有

$$\text{int } S=\{x\in R^n : g_j(x)<0, j=1,\cdots,p\}$$

设 $\text{int } S\neq\varnothing$，可构造 $B(x)$如下：

$$B(x)=\sum_{j=1}^{p}\frac{1}{|g_j(x)|^{\alpha}}, \alpha>0 \tag{13.29}$$

$$B(x)=-\sum_{j=1}^{p}\ln[-g_j(x)] \tag{13.30}$$

$$B(x)=\sum_{j=1}^{p}\frac{1}{\max\{0,-g_j(x)\}} \tag{13.31}$$

容易验证，(13.29)和(13.31)都满足前述关于 $B(x)$的要求. 对于(13.30)，$B(x)$不满足非负性，但由其所构成的增广目标函数能具有以上一切性质，这可从下段的假设条件和分析得到.

设

$$L\triangleq\max_{1\leqslant i\leqslant p}\sup_{x\in S}\{-g_j(x)\}<+\infty \tag{13.32}$$

再定义

$$\tilde{B}(x)=-\sum_{j=1}^{p}\ln\left[\frac{-g_j(x)}{L}\right] \tag{13.33}$$

则 $\tilde{B}(x)\geqslant 0, \forall x\in\text{int } S$，且其满足前述关于 $B(x)$的要求；比较(13.30)可知，由(13.33)所定义的 $\tilde{B}(x)$其对应于求解

$(\widetilde{\text{IP}})$ $$\min f(x)$$
$$\text{s.t. } g_j(x)/L\leqslant 0, j=1,\cdots,p$$

于是，(13.30)的可用性可从下面的关系看出：(IP)和$(\widetilde{\text{IP}})$的可行

集相同，最优解相同；它们各自的增广目标函数(分别关于(13.30)和(13.33)的)的最优解相同.

类似于§13.1的第2部分可研究内部罚函数的性质，且结论很相似，不赘述.

例13.10 $\min f(x)=x$, s.t. $1-x\leqslant 0, x\in R$.

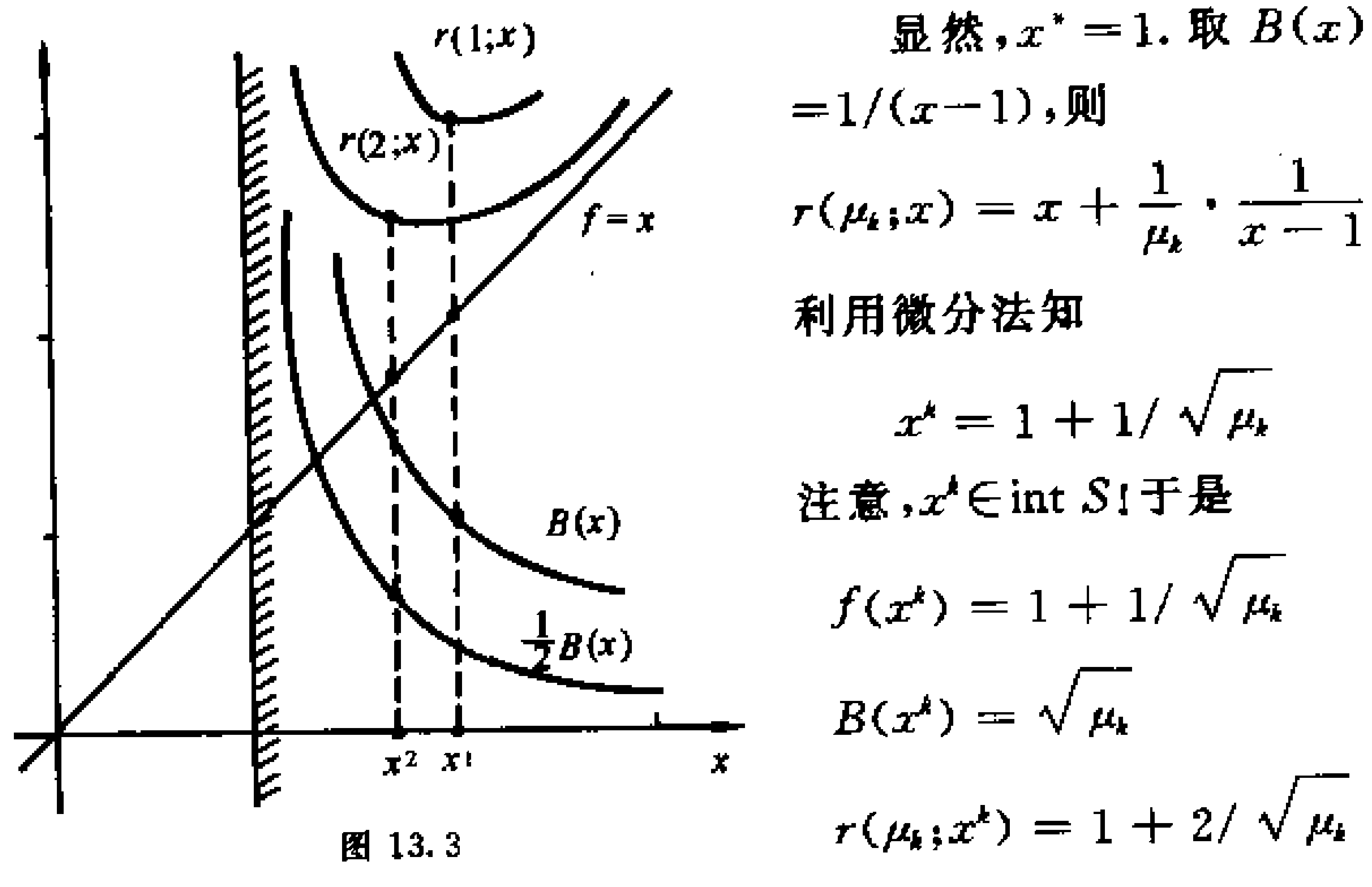


图 13.3

显然，$x^*=1$. 取 $B(x)=1/(x-1)$，则

$$r(\mu_k;x)=x+\frac{1}{\mu_k}\cdot\frac{1}{x-1}$$

利用微分法知

$$x^k=1+1/\sqrt{\mu_k}$$

注意，$x^k\in \text{int } S$！于是

$$f(x^k)=1+1/\sqrt{\mu_k}$$

$$B(x^k)=\sqrt{\mu_k}$$

$$r(\mu_k;x^k)=1+2/\sqrt{\mu_k}$$

据此可验证前面的结论. 读者可参见图13.3.

2. 内点的求法

内部罚函数法要求先找出一个内点. 这就提出了下面的问题. 设

$$\text{int } S=\{x\in R^n: g_j(x)<0, j=1,\cdots,p\}\neq\varnothing$$

求一个 $\bar{x}\in\text{int } S$.

值得一提的是，寻找内点的问题有广泛的应用.

利用内部罚函数的思想可求得 S 的一个内点.

算法13.11(内点求法之一)

步0　任取 $x^0\in R^n, \mu_1>0, k:=0$

步1　定出指标集 S_k 及 T_k：

$$S_k=\{j: g_j(x^k)\geqslant 0, j=1,\cdots,p\}$$

$$T_k=\{j:g_j(x^k)<0,j=1,\cdots,p\}$$

步2 若 $S_k=\varnothing$，则 $x^k\in\text{int }S$，停；否则

步3 以 x^k 为初始点，在保持对集合

$$R_k=\{x\in R^n:g_j(x)\leqslant 0,j\in T_k\}$$

可行性的情况下，求解

$$\min_{x\in\text{int }R_k} r_k(x)=\sum_{j\in S_k}g_j(x)-\frac{1}{\mu_{k+1}}\sum_{j\in T_k}\frac{1}{g_j(x)}$$

得最优解 $x^{k+1}\in\text{int }R_k$，其中

$$\text{int }R_k=\{x\in R^n:g_j(x)<0,j\in T_k\}$$

步4 取 $\mu_{k+2}>\mu_{k+1}$，$k:=k+1$，返步1.

上面的算法可在有限步内求解，具体地说，有

定理13.12 设 $\text{int }S\neq\varnothing$，设算法之步3总有解，设 $\mu_k\to+\infty$. 则算法必在有限步内终止于步2.

证 用反证法，设不存在 k_0，使得 $S_{k_0}=\varnothing$. 由于

$$\{1,\cdots,p\}\supseteq S_0\supseteq S_1\supseteq\cdots\supseteq S_k\supseteq S_{k+1}\supseteq\cdots$$

所以必存在 $k'(\geqslant 0)$，使得

$$\bar{S}=S_{k'}=S_{k'+1}=\cdots,\bar{S}\neq\varnothing \tag{13.34}$$

今取定 $y^0\in\text{int }S$，则当 $k\geqslant k'$ 时有

$$\sum_{j\in\bar{S}}g_j(x^k)-\frac{1}{\mu_k}\sum_{j\in\bar{T}}\frac{1}{g_j(x^k)}\leqslant\sum_{j\in\bar{S}}g_j(y^0)-\frac{1}{\mu_k}\sum_{j\in\bar{T}}\frac{1}{g_j(y^0)} \tag{13.35}$$

其中

$$\bar{T}=\{1,\cdots,p\}\backslash\bar{S}$$

由 $\bar{S}\neq\varnothing$，以及 $\mu_k\to+\infty$，以及 $y^0\in\text{int }S$，知，存在 $k''>k'$，使得

$$\sum_{j\in\bar{S}}g_j(y^0)-\frac{1}{\mu_{k''}}\sum_{j\in\bar{T}}\frac{1}{g_j(y^0)}<0 \tag{13.36}$$

比较(13.35)和(13.36)同时利用(13.34)知，对此 $k''>k'$ 有 $S_{k''}=\bar{S}$，从而

$$\sum_{j\in S_{k''}}g_j(x^{k''})-\frac{1}{\mu_{k''}}\sum_{j\in T_{k''}}\frac{1}{g_j(x^{k''})}<0 \tag{13.37}$$

另一方面，直接由 $S_{k'}$ 和 $T_{k'}$ 的定义知

$$\sum_{j\in S_{k'}} g_j(x^{k'}) - \frac{1}{\mu_{k'}}\sum_{j\in T_{k'}} \frac{1}{g_j(x^{k'})} \geqslant 0 \qquad (13.38)$$

(13.37)和(13.38)是矛盾的. 证毕.

同样，受外部罚函数的启发，我们也有

算法13.13(内点求法之二)

步0　任取 $r_0>1$，取定 $c>1$，$k:=0$；

步1　在 R^n 中极小化 $p_k(x)$，其中

$$p_k(x) = \sum_{j=1}^{p} [\max\{0, r_k+g_j(x)\}]^\alpha, \alpha>0\text{固定}$$

得最优解 $x^k\in R^n$；

步2　若 $p_k(x^k)=0$，则 $x^k\in \text{int } S$，停；否则

步3　$r_{k+1}:=r_k/c$，$k:=k+1$，返步1.

这个算法的有限步收敛性的证明很简单.

定理13.14　设 $\text{int } S\neq\varnothing$，设步1中 $\min\limits_{x\in R^n} p_k(x)$ 总有解，则算法13.13必在有限步内终止于步2.

证　存在 $r_k>0$，使得 $p_k(x^k)=0$

$\Leftrightarrow$存在 $r_k>0$ 及 x^k，使得 $\max\{0, r_k+g_j(x^k)\}=0$，$j=1,\cdots,p$

$\Leftrightarrow$存在 $r_k>0$ 及 x^k，使 $r_k+g_j(x^k)\leqslant 0$，$j=1,\cdots,p$

$\Leftrightarrow g_j(x^k)<0$，$j=1,\cdots,p$，即 $x^k\in \text{int } S$.

今任取定 $y^0\in \text{int } S$，记

$$\bar{\lambda} = \min_{1\leqslant j\leqslant p}\{-g_j(y^0)\} \qquad (13.39)$$

则 $\bar{\lambda}>0$. 于是可取 k，使其满足

$$0 < r_k = r_0/c^k \leqslant \bar{\lambda} \qquad (13.40)$$

于是显然有

$$r_k + g_j(y^0) \leqslant \bar{\lambda} + g_j(y^0) \leqslant 0, j = 1,\cdots,p \qquad (13.41)$$

从而就有

$$0 \leqslant \min_{x\in R^n} p_k(x) \leqslant p_k(y^0) \leqslant 0 \qquad (13.42)$$

这说明 $p_k(x^k)=0$. 证毕.

§ 13.3 恰当罚函数法

前面所述的两种罚函数法的缺点是本质上固有的，随着参数 μ_k 的增大，增广目标函数的 Hesse 阵会变成病态，数值计算便产生困难，因此导出那种只需参数的适当的值的条件或方法是可取的. 本节和下节就是解决这个问题的.

考虑如下的问题

$$(\mathrm{P}) \qquad \begin{aligned} &\min f(x) \\ &\text{s.t. } g_j(x) \leqslant 0, j=1,\cdots,p \\ &\qquad x \in C \end{aligned}$$

其中 $C \subseteq R^n$，f，g 都是 R^n 上的函数. 记(P)的最小值为 $v(P)$.

定义13.15 若存在 $\bar{\mu}_j \geqslant 0, j=1,\cdots,p$，使得

$$f(x) + \sum_{j=1}^{p} \bar{\mu}_j g_j(x) \geqslant v(P), \forall\, x \in C \qquad (13.43)$$

则称 $\bar{\mu} = (\bar{\mu}_1, \bar{\mu}_2, \cdots, \bar{\mu}_p)^T$ 为(P)的一个 Kuhn-Tucker 向量.

在第九章中，我们对(P)的 Kuhn-Tucker 向量的存在性作了较为详尽的讨论. 本节的结果可以看作那里的研究的应用.

对于(P)，定义一个函数 $E: C \to R$，其满足

$$\begin{cases} E(x) \geqslant 0, \forall\, x \in C \\ E(x) = 0 \text{ 当且仅当 } g_j(x) \leqslant 0, j = 1,\cdots,p \end{cases} \qquad (13.44)$$

这样的函数是存在的，例如可取 $E(x)$ 为

$$\sum_{j=1}^{p} \max\{0, g_j(x)\}, x \in C \qquad (13.45)$$

或

$$p \cdot \max\{0, g_1(x), \cdots, g_p(x)\}, \forall\, x \in C \qquad (13.46)$$

注意在 $C=R^n$ 时，(13.45)就是外部罚函数中的情况 I.

再考虑如下的单个极小化问题

$$(\mathrm{P}_\alpha) \qquad \min_{x \in C} \phi(x;\alpha) = f(x) + \alpha E(x)$$

定义13.16 若存在 $\alpha_0 \geqslant 0$，使得(P_{α_0})的解都是(P)的解，则说

$\phi(x;\alpha)$为(P)的恰当罚函数.

记(P_α)的最优值为 $v(P_\alpha)$.

引理13.17 (i)$v(P_\alpha)$关于 α 是单调上升的;

(ii)若 $\bar{x}$ 是($P_{\bar{\alpha}}$)的解且 $E(\bar{x})=0$,则 $\bar{x}$ 是(P)的解.

证 (i)$\forall\ \alpha_1<\alpha_2,\forall\ x\in C$,由(13.44)之第一式,

$$f(x)+\alpha_1E(x)\leqslant f(x)+\alpha_2E(x)$$

于是就有

$$\begin{aligned}v(P_{\alpha_1})&=\min_{x\in C}\phi(x;\alpha_1)\\&\leqslant\phi(x;\alpha_2),\forall\ x\in C\end{aligned}$$

由此即推得 $v(P_{\alpha_1})\leqslant v(P_{\alpha_2})$.

(ii)首先因为 $E(\bar{x})=0$,故由(13.44)之第二式及 $\bar{x}$ 为($P_{\bar{\alpha}}$)的解(这表明 $\bar{x}\in C$)可知 $\bar{x}$ 对(P)可行;再任取对(P)可行的 x,仍由(13.44)知 $E(x)=0$,从而

$$\begin{aligned}f(x)&=f(x)+\bar{\alpha}E(x)\\&\geqslant\min_{x\in C}f(x)+\bar{\alpha}E(x)\\&=f(\bar{x})+\bar{\alpha}E(\bar{x})\\&=f(\bar{x})\end{aligned}$$

证毕.

记(P_α)的最优解集为 X_α^*,记(P)的最优解集为 X^*. 下一个定理表明了恰当罚函数的存在性.

定理13.18 设(P_α)中的 $E(x)$由(13.45)或(13.46)给出,若(P)有一个 Kuhn-Tucker 向量,则存在 $\bar{\alpha}\geqslant0$,使得

$$X_\alpha^*=X^*,\forall\ \alpha\in(\bar{\alpha},+\infty)$$

证 设 $\bar{\mu}$ 是(P)的一个 Kuhn-Tucker 向量,则有

$$\begin{aligned}v(P)&\leqslant f(x)+\sum_{j=1}^{p}\bar{\mu}_jg_j(x)\\&\leqslant f(x)+\|\bar{\mu}\|_\infty E(x),\forall\ x\in C\end{aligned}$$

记 $\bar{\alpha}=\|\bar{\mu}\|_\infty$. 由上式结合引理13.17(i)便得

$$\begin{aligned}v(P)&\leqslant v(P_{\bar{\alpha}})\leqslant v(P_\alpha)\\&\leqslant\sup_{\alpha\geqslant0}\inf_{x\in C}\phi(x;\alpha)\end{aligned}$$

$$\leqslant \inf_{x\in C}\sup_{\alpha\geqslant 0}\phi(x;\alpha)$$

$$= v(P), \forall\ \alpha\in[\bar{\alpha}, +\infty) \tag{13.47}$$

上面的最后的等式是直接计算出来的.从而上面的不等式全成了等式,于是有

$$v(P)=v(P_{\bar{\alpha}})=v(P_\alpha), \forall\ \alpha\in[\bar{\alpha}, +\infty) \tag{13.48}$$

任取 $\hat{\alpha}>\bar{\alpha}$,设 $\hat{x}\in X_{\hat{\alpha}}^*$,利用(13.48)得

$$f(\hat{x})+\bar{\alpha}E(\hat{x})\geqslant v(P_{\bar{\alpha}})=v(P_{\hat{\alpha}})=f(\hat{x})+\hat{\alpha}E(\hat{x})$$

比较上式左右得 $E(\hat{x})\leqslant 0$,结合 $E(\hat{x})\geqslant 0$得 $E(\hat{x})=0$. 再结合 $\hat{x}\in X_{\hat{\alpha}}^*$ 利用引理13.17(ii)知 $\hat{x}\in X^*$,故有 $X_{\hat{\alpha}}^*\subseteq X^*$.

反之,设 $x^*\in X^*$. 先有

$$v(P)=\sup_{\alpha\geqslant 0}\phi(x^*;\alpha)\geqslant\phi(x^*;\hat{\alpha}) \tag{13.49}$$

另一方面又有

$$v(P_{\hat{\alpha}})=\inf_{x\in C}\phi(x;\hat{\alpha})\leqslant\phi(x^*;\hat{\alpha}) \tag{13.50}$$

结合(13.48)~(13.50)知上两式中的不等号实为等号,故

$$\inf_{x\in C}\phi(x;\hat{\alpha})=\phi(x^*;\hat{\alpha})$$

此即表明 $x^*\in X_{\hat{\alpha}}^*$. 故 $X^*\subseteq X_{\hat{\alpha}}^*$. 证毕.

注 若存在 $\alpha_2>\alpha_1\geqslant 0$使得

$$v(P_{\alpha_2})=v(P_{\alpha_1}) \tag{13.51}$$

则

$$v(P_{\alpha_1})=v(P_\alpha), \forall\ \alpha\in[\alpha_1, +\infty) \tag{13.52}$$

事实上,由引理13.17(i)及(13.51)即得

$$v(P_{\alpha_1})=v(P_\alpha), \forall\ \alpha\in[\alpha_1,\alpha_2]$$

今任取 $\alpha_3>\alpha_2$. 则存在 $\lambda\in(0,1)$使

$$\alpha_2=\lambda\alpha_1+(1-\lambda)\alpha_3$$

从而也有

$$\phi(x;\alpha_2)=\lambda\phi(x;\alpha_1)+(1-\lambda)\phi(x,\alpha_3), \forall\ x\in C$$

由此推得

$$v(P_{\alpha_2})\geqslant\lambda v(P_{\alpha_1})+(1-\lambda)v(P_{\alpha_3})$$

再结合(13.51)及 $\lambda\in(0,1)$ 由上式推得 $v(P_{\alpha_3})\leqslant v(P_{\alpha_2})$. 再由引理13.17(i)知 $v(P_{\alpha_3})\geqslant v(P_{\alpha_2})$, 从而 $v(P_{\alpha_3})=v(P_{\alpha_2})$.

根据上面的注及定理13.18可得如下的算法,它在有限步内终止于步2.

算法13.19(恰当罚函数法)

设(P)有一个 Kuhn-Tucker 向量,且设(P)的最优值有限. 设(P_α)中的 $E(x)$ 由(13.45)或(13.46)给出.

步0 任取 $c>1$, $\alpha_0>0$, $k:=0$;

步1 求解(P_{α_k});

步2 若 α_k 满足

$$v(P_{\alpha_k}) = v(P_{\alpha_{k-1}}) > -\infty \tag{13.53}$$

则

若(P_{α_k})有解 x^k, 则 x^k 为(P)的解;

若(P_{α_k})无解,则(P)无解.

两种情况都表明(P)的最优值为 $v(P_{\alpha_k})$. 停;否则,

步3 $\alpha_{k+1}:=c\cdot\alpha_k$, $k:=k+1$, 返步1.

对于一般情况的 $\phi(x;\alpha)$, 利用定理13.18证明中的部分思路可得如下的结论. 我们把证明留给读者.

定理13.20 设 $X^*\neq\varnothing$. 则 $\phi(x;\alpha)$ 为(P)的恰当罚函数的充要条件为,存在 $\hat{\alpha}\geqslant 0$ 以及 $\hat{x}\in C$ 使得

$$\phi(\hat{x};\alpha)\leqslant\phi(\hat{x};\hat{\alpha})\leqslant\phi(x;\hat{\alpha}),\forall\ \alpha\geqslant 0,\forall\ x\in C \tag{13.54}$$

注 这里关于恰当罚函数的大部分推导取自[29]. 其它有关结果可参见文献[5,13].

§13.4 乘子法

在本节中,我们恒假定所出现的函数在 R^n 上二次连续可微.

1. 等式约束的情况

考虑问题

(EP) $\min f(x)$

$\text{s.t. } h(x)=0$

其中 $x\in R^n, h\in R^m$. 设 x^* 是(EP)的满足二阶充分条件的解,即存在 $\lambda^*\in R^m$,它与 x^* 一起满足

$$h(x^*)=0 \tag{13.55a}$$

$$\nabla f(x^*)+\nabla h(x^*)\lambda^*=0 \tag{13.55b}$$

$$L(x^*)\text{ 在 } M(x^*)\text{ 上正定} \tag{13.55c}$$

其中

$$L(x^*)=\nabla^2 f(x^*)+\nabla^2 h(x^*)\lambda^*$$

$$M(x^*)=\{d\in R^n: \nabla h(x^*)^T d=0\}$$

$$\nabla h(x)=(\nabla h_1(x),\cdots,\nabla h_m(x))$$

$$\nabla^2 h(x)=(\nabla^2 h_1(x),\cdots,\nabla^2 h_m(x))$$

再考虑问题

(EP_{λ^*}) $\min f(x)+h(x)^T\lambda^*$

$\text{s.t. } h(x)=0$

注意,(EP_{λ^*})与(EP)的区别在于目标函数中多了一个其值为零的项 $h(x)^T\lambda^*$. 定义(EP_{λ^*})的增广目标函数为

$$H(x,\lambda^*,c)=f(x)+h(x)^T\lambda^*+\frac{c}{2}h(x)^T h(x) \tag{13.56}$$

其中 c 为实参数. H 在 x^* 处的梯度和 Hesse 阵分别为

$$\begin{aligned}\nabla_x H(x^*,\lambda^*,c)&=\nabla f(x^*)+\nabla h(x^*)\lambda^*+c\nabla h(x^*)h(x^*)\\&=\nabla f(x^*)+\nabla h(x^*)\lambda^*\\&=0\end{aligned} \tag{13.57a}$$

$$\nabla_{xx}^2 H(x^*,\lambda^*,c)=L(x^*)+c\nabla h(x^*)\nabla h(x^*)^T \tag{13.57b}$$

我们叙述(13.55c)和(13.57b)的一个关系.

引理13.21 设 $A\in R^{n\times n}, B\in R^{n\times m}$. 则对每个满足 $B^T d=0$ 的 $d\neq 0$ 都成立着 $d^T A d>0$ 的充要条件是:存在一个 $c^*>0$,使得对所有 $c\geqslant c^*$ 和 $d\neq 0$,有

$$d^T(A+cBB^T)d>0 \tag{13.58}$$

证 充分性显然. 下用反证法证必要性.

若存在$\{c_k\}$，$c_k\to\infty$，使得对应于每一个c_k，都有一个d^k，使得

$$(d^k)^T(A+c_kBB^T)d^k\leqslant 0 \tag{13.59}$$

不妨设$\|d^k\|=1$，则$\{d^k\}$有一个收敛子列，不妨设$d^k\to\overline{d}$. 于是就有$\|\overline{d}\|=1$. 今断言$B^T\overline{d}=0$. 不然的话，对充分大的k，(13.59)将不成立. 同时由(13.59)又得$(d^k)^TAd^k\leqslant 0$，从而又有

$$\overline{d}^TA\overline{d}\leqslant 0 \tag{13.60}$$

此式及$B^T\overline{d}=0$，$\|\overline{d}\|=1$便与必要性的假设条件矛盾. 证毕.

综上述，我们得到

定理13.22 若x^*与λ^*一起满足(EP)的局部极小的二阶充分条件，则存在$c^*\geqslant 0$，只要$c\geqslant c^*$，x^*就是由(13.56)所定义的$H(x,\lambda^*,c)$的在R^n上的满足二阶充分条件的局部极小点；反之，若存在x^0，λ^0，c^0，满足$h(x^0)=0$，以及x^0是无约束问题

$$\min_{x\in R^n} H(x,\lambda^0,c^0)$$

的局部极小点，则x^0也是(EP)的局部极小点.

在实际计算时，λ^*和c_0是不知道的. 假定我们已估计出一个c_0，希望λ^*由一些迭代得到. 先给定λ^k，求解$\min\limits_{x\in R^n}H(x,\lambda^k,c_0)$，得解$x^k$. 由一阶必要条件得

$$\begin{aligned}0&=\nabla_xH(x^k,\lambda^k,c_0)\\&=\nabla f(x^k)+\nabla h(x^k)[\lambda^k+c_0\cdot h(x^k)]\end{aligned} \tag{13.61}$$

下一步的迭代则求解$\min\limits_{x\in R^n} H(x,\lambda^{k+1},c_0)$，其中

$$\lambda^{k+1}=\lambda^k+c_0\cdot h(x^k) \tag{13.62}$$

一个合适的算法应考虑到c_0增加的可能性，以及x^k满足约束$h(x)=0$的程度. 我们描述之.

算法13.23(乘子法)

步0 任选$\lambda^0\in R^m$，$c_0>0$. 给定允许误差$\varepsilon>0$. $k:=0$；

步1 求解$\min\limits_{x\in R^n} H(x,\lambda^k,c_k)$，得解$x^k$；

步2 若$\|h(x^k)\|\leqslant\varepsilon$，停，$x^k$为(EP)之近似解；否则

步3 若$k=0$，则$\lambda^{k+1}:=\lambda^k+c_kh(x^k)$，$k:=k+1$，返步1；

若 $k\geqslant 1$ 且 $\|h(x^k)\|\geqslant\|h(x^{k-1})\|$，则增大 c_k（例如 $c_k:=10c_k$），重解步1中之规划；

若 $k\geqslant 1$ 且 $\|h(x^k)\|<\|h(x^{k-1})\|$，则 $\lambda^{k+1}:=\lambda^k+c_kh(x^k)$，$k:=k+1$，返步1.

这个算法具有较好的性质. 详论从略.

2. 含有不等式约束的情况

将上节的结果推广到本节只需要一点简单的技巧. 并且，从以下的推导来看，我们仅须考虑只含有不等式约束的情况. 于是，考虑问题

(IP)
$$\min f(x)$$
$$\text{s.t. } g_j(x)\leqslant 0,\ j=1,2,\cdots,p$$

引进新变量

$$z=(z_1,\cdots,z_p)^T\in R^p$$

可将(IP)化为等价的

(IP)′
$$\min_{x,z} f(x)$$
$$\text{s.t. } g_j(x)+(z_j)^2=0,\ j=1,\cdots,p$$

其对应的增广 Lagrange 函数为

$$\tilde{G}(x,z,\mu,c)=f(x)+\sum_{j=1}^{p}\mu_j[g_j(x)+(z_j)^2]$$
$$+\frac{c}{2}\sum_{j=1}^{p}[g_j(x)+(z_j)^2]^2 \tag{13.63}$$

$\tilde{G}$ 关于 x,z 取极小应满足一阶必要条件，特别地应有

$$0=\frac{\partial\tilde{G}}{\partial z_j}$$
$$=2\mu_jz_j+2c[g_j(x)+(z_j)^2]z_j,\ j=1,\cdots,p \tag{13.64}$$

解之，得

$$z_j=0 \tag{13.65a}$$

或

$$\mu_j+c[g_j(x)+(z_j)^2]=0 \tag{13.65b}$$

由(13.65b)解得

$$(z_j)^2 = -\frac{1}{c}[\mu_j + cg_j(x)] \tag{13.66}$$

结合(13.65a)和(13.66)，$(z_j)^2$的一个合适的取法为

$$(z_j)^2 = \frac{1}{c}\max\{0, -[\mu_j + cg_j(x)]\} \tag{13.67}$$

将(13.67)代入(13.63)以消去 z，得

$$G(x,\mu,c) = f(x) + \frac{1}{2c}\sum_{j=1}^{p}\{[\max\{0,\mu_j + cg_j(x)\}]^2 - \mu_j^2\} \tag{13.68}$$

仿(13.62)，相应的乘子迭代式可以计算得

$$\begin{aligned}\mu_j^{k+1} &= \mu_j^k + c[g_j(x^k) + (z_j)^2] \\ &= \mu_j^k + c\left[g_j(x^k) + \frac{1}{c}\max\{0, -[\mu_j^k + cg_j(x^k)]\}\right] \\ &= \max\{0, \mu_j^k + cg_j(x^k)\}, j = 1,\cdots,p \end{aligned} \tag{13.69}$$

(13.68)和(13.69)就是对含不等式约束的规划(IP)构造乘子法算法的基础，读者可仿算法13.23完成相应的算法.

习　题

1. 对§13.1中所定义的(P)和 $q(\mu_k;x)$，证明，若存在 μ_{k_0}使得

$$\min_{x\in R^n} q(\mu_{k_0};x)$$

的解 $x^{k_0}\in S$，则对任意 $\mu_k \geq \mu_{k_0}$，$\min\limits_{x\in R^n} q(\mu_k;x)$的解都是(P)的解.

2. 证明推论13.5.

3. 证明引理13.7.

4. 证明推论13.9.

5. 试用对数型内部罚函数法求解

$$\begin{aligned}&\min\ x_1+2x_2 \\ &\text{s.t.}\ -(x_1)^2+(x_2)^2\geq 0.\end{aligned}$$

6. 对问题

$$\min\ x, \text{s.t.}\ 1-x\leq 0, x\in R$$

验证定理13.18.

7. 用乘子法求解

$$\min \frac{1}{3}(x_1+1)^3+x_2$$

$$\text{s.t. } x_1-1\geqslant 0, x_2\geqslant 0$$

8. 对不等式约束问题

(IP) $$\min f(x)$$

$$\text{s.t. } g_j(x)\leqslant 0, j=1,\cdots,p$$

通过引进变量 $z=(z_1,\cdots,z_p)^T\in R^p$，可将其化为等价的

(IP)′ $$\min_{x,z} f(x)$$

$$\text{s.t. } g_j(x)-z_j=0$$

$$z_j\leqslant 0$$

$$j=1,\cdots,p$$

先写出关于(IP)′的只含有等式约束的增广阵，再写出此增广阵在约束条件 $z_j\leqslant 0, j=1,\cdots,p$ 下关于 z 的极小化的一阶必要条件，由此导出(13.68)和(13.69).

9. 对 $\min_{x\in S} f(x)$，定义 $q(\mu;x)=f(x)+\mu P(x)$，其中 $P(x)$ 如§13.1中所述，f 在 R^n 中连续. 选一个 $\epsilon>0$，取单调增加趋于 $+\infty$ 的 $\{\mu_k\}$，令 x^k 是满足

$$q(\mu_k;x^k)\leqslant \min_{x\in R^n} q(\mu_k;x)+\epsilon \tag{13.70}$$

的一个 R^n 中的点. 证明，$\{x^k\}$ 的任一聚点 $\bar{x}$ 满足 $\bar{x}\in S$，且

$$f(\bar{x})\leqslant \min_{x\in S} f(x)+\epsilon \tag{13.71}$$

10. 外部罚函数法和内部罚函数法可结合使用. 对问题

(MP) $$\min f(x)$$

$$\text{s.t. } x\in S\cap T$$

假定 $P(x)$ 对 S，$B(x)$ 对 T 分别满足§13.1和§13.2中的有关条件. T 是闭集，且 S 和 T 一起满足

$$S\cap T=\text{cl}(S\cap \text{int } T) \tag{13.72}$$

考虑

$$M(\mu;x)=f(x)+\mu P(x)+\frac{1}{\mu}B(x) \tag{13.73}$$

令 $\{\mu_k\}$ 单调增加，且 $\mu_k\to+\infty$. 令 x^k 是

$$\min_{x\in \text{int } T} M(\mu_k;x) \tag{13.74}$$

的解. 若 f 在 R^n 上连续，则

(i) $\{x^k\}$ 的任一聚点都是(MP)的解；

(ii) 设 x^* 为 $\{x^k\}$ 的聚点，则

$$\lim_{k\to\infty} M(\mu_k; x^k) = f(x^*)$$

$$\lim_{k\to\infty} \mu_k P(x^k) = 0$$

$$\lim_{k\to\infty} \frac{1}{\mu_k} B(x^k) = 0$$

11. 考虑问题

$$(\mathrm{P}_H) \qquad \min f(x) \quad \text{s.t. } g(x) \in H, x \in C$$

其中 $C \subseteq R^n$，$g \in R^p$，H 为 R^p 中预先给定的闭集. 定义

$$\phi(\alpha; x) = f(x) + \alpha\rho(g(x), H) \tag{13.75}$$

其中 ρ 为 R^p 中的某个距离函数.

(i) 证明，若 (P_H) 的解集 X^* 非空，则 $\phi(\alpha; x)$ 为 (P_H) 的恰当罚函数的充要条件是，存在 $\bar{x} \in C$，$\bar{\alpha} \geq 0$，使得

$$\phi(\alpha; \bar{x}) \leqslant \phi(\bar{\alpha}; \bar{x}) \leqslant \phi(\bar{\alpha}; x), \forall\, \alpha \geqslant 0, \forall\, x \in C \tag{13.76}$$

此时

$$X^* = X_\alpha^*, \forall\, \alpha \in (\bar{\alpha}, +\infty) \tag{13.77}$$

其中 X_α^* 表 $\min\limits_{x \in C} \phi(\alpha; x)$ 的解集合，且所涉及的解都是全局性的；

(ii) 对 § 13.3中的(P)，将它改写为 (P_H) 的形式；

(iii) 指定 p 维向量空间中的一个距离 ρ，对 § 13.3中的(P)写出具体的 $\rho(g(x), H)$.

注 关于(13.75)中的 $\phi(\alpha; x)$ 为 (P_H) 的恰当罚函数的另一个充要条件可参见[5].

参 考 文 献

[1] Avriel, M., Nonlinear Programming: Analysis and Methods, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1976.

[2] Bazaraa, M. S. and Shetty, C. M., Nonlinear Programming: Theory and Algorithms, John Wiley & Sons, Inc., New York, 1979.

[3] Clarke, F. H., Generalized gradients and applications, Transaction of the American Mathematical Society, Vol. 205, 247～262, 1975.

[4] Dantzig, G. B., Linear Programming and Extensions, Princeton University Press, Princeton, New Jersey, 1963.

[5] Dien, P. H., Mastroeni, G., Pappalardo, M. and Quang, P. H., Regularity conditions for constrained extremum problem via image space, JOTA, Vol. 80, 19～37, 1994.

[6] Hansen, C. T., Madsen, K. and Nielsen, H. B., Optimization of pipe networks, Mathematical Programming, Vol. 52, 45～58, 1991.

[7] Hanson, M. A., On sufficiency of the Kuhn-Tucker conditions, JMAA, Vol. 80, 545～550, 1981.

[8] Karmarkar, N., A new polynomial-time algorithm for linear programming, Combinatoria, Vol. 4, 373～395, 1984.

[9] Khachian, L. G., A polynomial algorithm in linear programming, Soviet Mathematics, Doklady, Vol. 20, 191～194, 1979.

[10] Kuhn, H. W. and Tucker, W. A., Nonlinear Programming, Proceedings 2nd Berkeley Symposium on Mathematical Statistics and Probability, J. Neyman (Ed.), University of California Press, Berkeley, California, 1951.

[11] Luenberger, D. G., Introduction to Linear and Nonlinear Programming, Addison-Wesley, Reading, Massachusetts, 1973.

[12] Mangasarian, O. L., Nonlinear Programming, McGraw-Hill Book Co., New York, 1969.

[13] Mangasarian, O. L., Sufficiency of exact penalty minimization, SIAM J. Control and Optimization, Vol. 23, 30～37, 1985.

[14] Martin, D. H., The essence of invexity, JOTA, Vol. 47, 65～76, 1985.

[15] Rajasekera, J. R. and Fang, S. C., Deriving an unconstrained convex program for linear programming, JOTA. Vol. 75, 603～612, 1992.

[16] Roberts, A. W. and Varbero, D. E., Another proof that convex functions are locally Lipschitz, American Mathematical Monthly, Vol. 81, 1014～1016, 1974.
[17] Rockafellar, R. T., Convex Analysis, Princeton University Press, Princeton, New Jersey, 1972.
[18] Xu, Z. K., Duality in generalized nonlinear fractional programming, JMAA, Vol. 169, 1～9, 1992.
[19] Xu, Z. K., On invexity-type nonlinear programming problems, JOTA, Vol. 80, 135～148, 1994.
[20] Xu, Z. K., Generalization of nonhomogeneous Farkas' lemma and applications, JMAA, Vol. 186, 726～734, 1994.
[21] Xu, Z. K. and Fang, S. C., Unconstrained convex programming approach to linear programming, JOTA, Vol. 86, 745～752, 1995.
[22] Xu, Z. K., Local saddle points and convexification for nonconvex optimization problems, JOTA, Vol. 94, 739～746, 1997.
[23] 邓乃扬，无约束最优化计算方法，科学出版社，1982.
[24] 赵凤治，线性规划计算方法，科学出版社，1981.
[25] 管梅谷，郑汉鼎，线性规划，山东科学技术出版社，1983.
[26] 徐增堃，带约束的非线性 l_1 问题，高等学校计算数学学报，Vol. 10, 193～201, 1988.
[27] 徐增堃，约束 TRA 方法的全局收敛性，浙江师范大学学报，Vol. 15, 19～23, 1992.
[28] 徐增堃，Rockafeller 的一个择一定理的简证，浙江师范大学学报，Vol. 16, 12～14, 1993.
[29] 徐增堃，关于恰当罚函数的注记，浙江师范大学学报，Vol. 18, 1～4, 1995.
[30] 徐增堃，双线性 Stackelberg 问题，浙江师范大学学报，Vol. 19, 16～20, 1996.

现代数学基础丛书

